A Gricean Account of Assertion: closing the gap between the philosophy of language and the epistemology of testimony

David J. J. Austin

Doctor of Philosophy

University of York

Philosophy

December 2020

Abstract

Assertion plays a central role within the philosophy of language and the epistemology of testimony. However, the literature on assertion within the philosophy of language has developed from interests in a variety of different language-based issues, by means of a variety of different apparent cases of the speech act. And, largely detached from that, the literature on assertion within the epistemology of testimony has developed from interests in a variety of different epistemological issues, again, by means of a variety of different apparent cases of the speech act. The lack of integration between these two areas of philosophy on the topic of assertion has resulted in uncertainty about the concept, which has led some to think that it is not a concept we need to explain any significant component of human behaviour. Although I sympathise with this pessimistic view, ultimately, I think it is incorrect. Rather, philosophy simply has not made up its mind about how the concept of assertion is best understood, and so what is needed is a proposal for how it might fruitfully be applied. This involves reflecting on what philosophical work the concept of assertion should do, stipulating a definition based on these reflections, and then developing a theoretical elaboration which is guided by, connected to, and framed within broader philosophical theories, which can carry out said philosophical work. In doing this, and by drawing on Grice's framework of communication and the notion of knowledge transmission specifically, I offer a novel Gricean account of assertion. My aim is not only to reaffirm the concept of assertion, but also to provide innovative solutions to problems pertaining to assertion's role in communication and the spread of knowledge, as well as offer principled verdicts on whether certain apparent cases of assertion really are best understood as cases of assertion.

Declaration

I declare that this thesis is a presentation of original work and that I am the sole author. This work has not previously been presented for an award at this, or any other, university. All sources are acknowledged as references.

Acknowledgments

My greatest debt in completing this thesis is owed to my supervisors, Greg Currie and Tom Stoneham, whose philosophical insight and guidance has been invaluable.

I would also like to thank my thesis advisory panel, Chris Jay and Peter Lamarque, for always being on call.

In spring 2019, I visited my old haunt, the University of Hertfordshire, where I worked under the supervision of Craig Bourne. He holds a special place in my philosophical life for his continuous inspiration, encouragement, and good humour, for which I am forever thankful.

My doctoral research was kindly funded by the academic journal *Mind & Language*, through which I came to know a whole team of distinguished academics, but, more importantly, wonderful people. Those who deserve special mention for going out of their way to comment on draft chapters of this thesis are Robyn Carston, Nat Hansen, Guy Longworth, Louise Richardson, and Deirdre Wilson.

I have also benefitted greatly from conversations with Pål Antonsen, Emily Caddick Bourne, Manuel García-Carpintero, Mitch Green, Katherine Hawley, Richard Moran, Mark Pinder, Andreas Stokke, Timothy Williamson, and Stephen Wright.

Finally, I wish to thank my family for their unyielding support and generosity: my parents, Catriona and Gerry; my brother and his wife, Matthew and Rachel; my sister and her fiancé, Hannah and Gary; and last, but certainly not least, my fiancée, Frances.

David J. J. Austin

York

December 2020

Table of Contents

Answers do not matter so much as questions, said the Good Fairy. A good question is very hard to answer. The better the question the harder the answer. There is no answer at all to a very good question.

—Flann O'Brien, At Swim-Two-Birds

1. Introduction

Assertion is troublesome. On one hand, you will find that talk of it is ubiquitous in philosophy. On the other hand, you cannot always be sure what precisely you have found talk of, or indeed whether you have found talk of anything concrete at all. I have been dogged by this ever since I was an undergraduate student working on the liar paradox. (As I saw it then, and as I still see it now, sentences of the form, 'This sentence is false', are not 'assertable'.) And it continued into my masters when I was working on the nature of lying. (As I saw it then, and as I still see it now, lying is typically, though not necessarily, a matter of a speaker 'asserting' something which they take to be false.) By the time I was ready to undertake doctoral study, I found myself inexorably drawn to – and eventually embroiled in – the debate on the nature of assertion. However, I was dissatisfied with the state of the debate. It seemed to me that philosophers were often talking past each other and getting caught up in something which was gradually losing significance in the eyes of spectators. Yet, it also seemed to me that assertion was undoubtedly worthy of philosophical attention, with many philosophically interesting and important things having been said about it, being said about, and indeed still needing to be said about it. This thesis therefore may be seen as a reaction to the frustration I have felt working with, on, and sometimes against assertion. More specifically, though, this thesis is my attempt to help determine not merely what we *do* talk about when we talk about assertion, but what we *should* talk about when we talk about assertion. Here is a brief overview.

In *Assertion & Explication*, I introduce the philosophical term of art 'assertion' as it is understood in the philosophy of language and the epistemology of testimony – its main philosophical areas. This includes providing an outline of the ongoing debate about the nature of assertion in the philosophy of language and the important though sometimes unclear status assertion has in the epistemology of testimony. I explain that the way in which philosophical discussion about assertion has materialised, especially the lack of integration between the philosophy of language and the epistemology of testimony on the topic, has led to doubts about the effectiveness of the concept. Drawing on Carnap's (1947;

1950) notion of 'explication', however, I hold that philosophy simply has not made up its mind about how the concept of assertion is best understood, and that what is needed is a proposal for how it might fruitfully be applied. Specifically, I propose that assertion is best understood simply in terms of the paradigm case. In Carnap's terminology, this is our candidate 'explicatum'. I acknowledge the potential worry that restricting assertion in the way that I suggest is implausible, given the apparent non-paradigmatic (or borderline) cases of assertion that exist in the philosophy literature. However, I plead that our inquiry into the concept should not be guided by such cases. Instead, I recommend that our inquiry should be guided by, connected to, and framed within broader philosophical theories, especially from the philosophy of language and the epistemology of testimony. Essentially, my idea is that we are best placed to demonstrate the effectiveness of the concept of assertion only when our candidate explicatum has received a full theoretical elaboration.

In *Putting the Epistemology of Testimony to Work*, I call upon the epistemology of testimony to consider the ways in which the speech act of assertion can put an individual in a position to know the truth of what is asserted. I explain that such considerations are apt to guide us in our inquiry and thus lay the groundwork for constructing a theoretical elaboration of our candidate explicatum. Although I show that there are various ways in which assertion can put an individual in a position to know the truth of what is asserted, I pick out one particular way in which assertion can put an individual in a position to know the truth of what is asserted which I think illuminates the speech act as I understand it: 'knowledge transmission'. To fully understand this independently plausible notion from the epistemology of testimony literature, however, we must understand the nature of the interpersonal relation at play between speakers and hearers in such cases. And, to understand this, we must understand the conversational dynamics which underpin communicative exchanges generally. I exhibit the current confusion in the epistemology of testimony literature about how to understand these things, the result of which is that I find myself in a predicament. I intend to construct a theoretical elaboration of our candidate explicatum in light of the notion of knowledge transmission. But I do not have an accurate account of the conversational dynamics which underpin communicative exchanges generally. As such, I do not have an accurate account of the nature of the interpersonal

relation at play in cases of knowledge transmission specifically. This, in turn, means that Ido not have a fully developed view of knowledge transmission in light of which I can construct a theoretical elaboration of our candidate explicatum.

In *A Gricean Account of Assertion,* I use, revise, and develop Grice's (1957; 1969; 1975; 1989) framework of communication from the philosophy of language literature to construct an account of the conversational dynamics which underpin communicative exchanges generally. I then use this to construct an account of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, provides us with a fully developed view of knowledge transmission in light of which I construct my theoretical elaboration of our candidate explicatum.

In *Exploring the Exotica*, I defend my theoretical elaboration of our candidate explicatum – my Gricean account of assertion – against some potential objections.

2. Assertion & Explication

In this chapter, I introduce the philosophical term of art 'assertion' as it is understood in the philosophy of language and the epistemology of testimony – its main philosophical areas. What we find, however, is not only that there is ongoing debate about what assertion is within these respective philosophical areas, but that there is also a lack of integration between these two philosophical areas on the topic of assertion. Unfortunately, this has resulted in uncertainty about the concept of assertion, which has led some to think that it is not a concept we need to explain any significant component of human behaviour. Although I sympathise with this pessimistic view, ultimately, I think it is incorrect. Drawing on Carnap's (1947; 1950) notion of explication, I hold that philosophy simply has not made up its mind about how the concept of assertion is best understood, and that what is needed is a proposal for how it might fruitfully be applied. This involves reflecting on what philosophical work the concept of assertion should do, stipulating a definition based on these reflections, and then developing a theoretical elaboration which can carry out said philosophical work. By reflecting on the least contentious case of assertion in the philosophy literature, i.e., the paradigm case, I propose that assertion is best understood simply in terms of this. Although a restricted conception of assertion of this sort might seem unpalatable given the various apparent borderline cases of assertion across the philosophy of language and the epistemology of testimony literature, I plead that our inquiry into the concept of assertion should not be guided so hastily and wholeheartedly by such things. Indeed, I argue that given the uncertainty about the concept of assertion (outside of the paradigm case) it is unwise to put much confidence in such apparent borderline cases of the speech act. I recommend instead that our inquiry into the concept should be guided by, connected to, and framed within broader philosophical theories, especially from the philosophy of language and the epistemology of testimony. As such, confidence about whether apparent borderline cases of assertion are best understood as cases of assertion may come more readily from the theoretical elaboration of our stipulated definition of the concept.

2.1. **Assertion in the Philosophy of Language**

In the philosophy of language, 'assertion' has become a term of art, which is intended to pick out a certain kind of act, namely, a type of speech act. To understand what a speech act is, it is helpful to consider Austin's (1962) distinction between a 'locutionary act' and an 'illocutionary act'. According to Austin, a locutionary act is:

> the utterance of certain noises, the utterance of certain words in a certain construction and the utterance of them with a certain "meaning" in the favourite philosophical sense of that word, i.e., with a certain sense and with a certain reference. The act of "saying something" in this full normal sense I call, i.e., dub, the performance of a locutionary act. (94-95)

A locutionary act (or saying), then, essentially is the expression of a proposition. Cappelen (2011) offers a helpful illustration here. He starts by inviting us to consider the following sentence:

> *N*: There are naked mole-rats in Sweden. (23)

As he observes, this sentence can be uttered by an English-speaking person to express the proposition that *there are naked mole-rats in Sweden*. Call this proposition '*n*', which is the conventionally fixed meaning of *N* in English.[1] (Of course, an English-speaking person need not utter *N* in order to express *n*, since *n* can be expressed using other English sentences, as well as non-English sentences.) Cappelen points out further that when a speaker expresses a proposition something that they can subsequently do is refer to it with a demonstrative: 'That is the proposition I have been asked to express'. Moreover, they can refer to it

[1] There is some ambiguity here regarding the phrase 'uttered sentence'. For example, some may wish to include under this label smoke signals or car horns, whereas others may not. For the purposes of this thesis, I am liberal in my use of 'uttered sentence'.

anaphorically: 'It is an interesting thought, I have no idea whether it is true or false, I have no evidence either way, and I did not mean to convey to anyone that I think it is true. I just wanted to express it'. Saying something, then, may be understood as the linguistic analogue of attending to a thought. We can contrast this, however, with someone who makes an utterance but does not know the language of which the sentence is part. Here, Cappelen invites us to consider the following sentence:

*N**: Det er mange svensker som jobber i Oslo. (Ibid.)

As he observes, although this sentence can be used by a Norwegian-speaking person to say that *there are many Swedes working in Oslo*, it cannot be used by those who do not speak Norwegian to say this, even though such a speaker can nonetheless utter the same sentence in just the same way. Why? Because, as Austin says, to perform a locution a speaker must make an utterance with a certain 'meaning' in the favourite philosophical sense of that word (i.e., with a certain sense and with a certain reference), which, in this instance, they cannot. Further details would be needed in order to give a full account of saying, but the following working definition will suffice for our purposes:

Austinian saying: *A* says that *p* iff:

(i) Sentence *S* conventionally means that *p* in language *L* among *L*-utterers
(ii) *A*, as an *L*-utterer, utters sentence *S*
(iii) *A* intends to express *p* by uttering sentence *S*.[2]

Having specified what a locutionary act is, Austin characterises the notion of an illocutionary act ostensively, giving examples of the following sort: making an assertion, asking a question, issuing an apology, giving an order, inviting a supposition, delivering a reminder, offering up a speculation, etc. Importantly, on Austin's picture, the performance

[2] I am sympathetic to Lewis's (1969; 1975) account of conventions of language.

of an illocutionary act involves the performance of a locutionary act. That is, speakers perform illocutions by means of performing locutions.[3] A speech act, then, may be understood as an illocutionary act, where an illocutionary act itself is built on top of a locutionary act.[4] A speaker, therefore, may perform a locution (or saying) in the service of performing one of any number of illocutions (or speech acts).

Very roughly, assertion is typically understood as a type of speech act in which something is claimed to hold. For example, a speaker might assert, with respect to some particular cat and some particular mat within some particular context, that *the cat is on the mat.* But what precisely makes an assertion the type of speech act that it is, rather than some other type of speech act? Put differently, since only some sayings are assertions, what is it that singles out this sub-set of sayings as assertions?[5] The answer to this question is still outstanding in the philosophy of language literature. There are, however, four broad accounts on the market.[6] Adopting Goldberg's (2015) terminology, I call these the 'attitudinal account', the 'common

---

[3] Although it is not unanimous, it is at least generally held that asserting that *p* involves saying that *p*. For the purposes of this thesis, I do not contest this generally held position. But for an argument against this, see Viebahn (2017). Of course, there are various questions and issues that arise regarding the notion of saying and its involvement in the speech act of assertion, but those questions and issues are ones that all parties within the debate have a shared interest in answering and solving. As such, whatever those answers and solutions are, they do not adjudicate on the assertion debate.

[4] As McFarlane (2011: 80) observes, we may use the term 'proposition' for the content of locutions, but the content of a locution used in the service of the illocutionary act of asking a question is perhaps best understood as a set of propositions (the possible answers to the question), not a proposition. We need not be held up by this detail, though. See Hamblin (1973), Karttunen (1977), and Groenendijk & Stokhof (1997).

[5] Typically, the focus is on the saying of declarative sentences. Williamson (2000: 258), for example, says, '[i]n natural language, the default use of declarative sentences is to make assertions'. However, I do not address this here.

[6] McFarlane (2011) was first to identify and taxonomize the four broad accounts of assertion within the philosophy of language literature, and so §2.1.1, §2.1.2, §2.1.3, and §2.1.4, which follow his taxonimisation and his general presentation of the four broad accounts of assertion, are deeply indebted to him..

ground account', the 'commitment account', and the 'constitutive rule (or norm) account', respectively.[7] I introduce them briefly below.[8]

2.1.1. **The Attitudinal Account**

The attitudinal account is developed by Bach & Harnish (1979), who write:

> In uttering *e*, utterer *U* asserts that *p* if *U* expresses:
>
> (i) the belief that *p*
> (ii) the intention that hearer *H* believe that *p*. (42)

According to Bach & Harnish, the notion of expressing an attitude is best understood in terms of what they call 'R-intentions', where R-intentions are intentions to bring something about in a hearer by way of the hearer's recognition of this very intention.[9] (Note, the notion of 'R-intentions' is inspired by Grice's (1957; 1969; 1989) theory of meaning, or 'non-natural meaning'.) Bach & Harnish unpack the notion of expressing an attitude as follows: for utterer *U* to express an attitude is for *U* to R-intend hearer *H* to take *U*'s utterance as a reason to think that *U* has that attitude.[10] On this view, then, what makes an assertion the type of speech act that it is, rather than some other type of speech act, is that a speaker says

---

[7] Although my taxonimisation and general presentation of the four broad accounts of assertion in §2.1.1, §2.1.2, §2.1.3, and §2.1.4 follow McFarlane's (2011), my exposition of the four broad accounts of assertion in §2.1.1, §2.1.2, §2.1.3, and §2.1.4 are deeply indebted to Goldberg (2015), who provides helpful, updated re-articulations of McFarlane's earlier expositions.

[8] For the purposes of this thesis, I present each of these accounts of assertion as though they involve the notion of saying. However, some of the original iterations, and indeed some newly developed iterations, do not explicitly state this, and some may even deny it. However, as expressed in fn.3, it is at least generally held that the act of asserting that *p* involves the act of saying that *p*, so my presentation of these four broad accounts as involving the notion of saying is not problematic.

[9] See Goldberg (2015: 10).

[10] See Bach & Harnish (1979: 15).

that *p* and intends that their utterance be regarded by the hearer as a reason both to think that the speaker believes that *p* and that the speaker intends the hearer to believe that *p*.

2.1.2. **The Common Ground Account**

The common ground account is developed by Stalnaker (1978; 1999; 2002), who writes:

> [T]he essential effect of an assertion is to change the presuppositions of the participants in a conversation by adding the content of what is asserted to what is presupposed. (Stalnaker 1978: 86)

According to Stalnaker, conversations take place and evolve against a background of mutually shared information, which is established by the presuppositions of the participants of the conversation. He calls this the 'common ground'. He defines the common ground in terms of 'acceptance', which, he claims, is a non-factive propositional attitude that is weaker than belief. He employs the notion of acceptance rather than belief because he maintains that successful communication is compatible with presuppositions that are mutually recognised as false, where the information being presupposed is nonetheless actually available, and not just assumed or pretended to be available. He writes:

> Common ground: it is common ground that *p* in a group if all members accept (for the purpose of the conversation) that *p*, and all believe that all accept that *p*, and all believe that all believe that all accept that *p*, etc. (Stalnaker 2002: 716)

On this view, what makes an assertion the type of speech act that it is, rather than some other type of speech act, is that a speaker says that *p* and thereby proposes that *p* become part of the common ground.

2.1.3. **The Commitment Account**

The commitment account is introduced by Peirce (1934), who writes:

> [T]o assert a proposition is to make oneself responsible for its truth. (384)

Like Stalnaker's approach, the commitment account defines assertion in terms of its 'essential effect'.[11] However, it regards the essential effect as the conferral of new commitments or obligations upon the speaker. The notion of a speaker conferring new commitments or obligations upon themselves may be fleshed out in various ways. For instance, Brandom (1983; 1994) maintains that when a speaker makes an assertion that *p*, they undertake a commitment to vindicate their entitlement to *p* when challenged, which, in turn, entitles others to assert that *p* on the speaker's authority. The notion of a speaker conferring new commitments or obligations upon themselves may be understood in epistemic terms. Whatever variation it takes, though, it suffices to say that, on this view, what makes an assertion the type of speech act that it is, rather than some other type of speech act, is that the speaker says that *p* and confers a commitment upon themselves to the truth of *p*, which involves incurring the obligation to vindicate *p*.

2.1.4. **The Constitutive Rule (or Norm) Account**

The constitutive rule (or norm) account is developed by Williamson (1996; 2000), who writes:

> One might suppose […] that someone who knowingly asserts a falsehood has thereby broken a rule of assertion, much as if he had broken a rule of a game; he has cheated. (1996: 489)

According to Williamson, if we assume that any respect in which the performance of an act can deserve praise or criticism is a rule (or norm) for that act, then the speech act of assertion has many rules (or norms). He claims, for instance, that assertions may be praised as true, informative, relevant, well-phrased, or polite, and they may be criticised as false,

[11] See Goldberg (2015: 11).

uninformative, irrelevant, ill-phrased, or rude. However, he holds that not all rules (or norms) for assertion are on a par with each other, and that a certain rule (or norm) is more intimately connected to the speech act of assertion than others, namely:

> (KA) One must: assert that *p* only if one knows that *p*. (494)

To illustrate this idea, he attempts to draw an analogy between assertion and the game of chess. He claims that the rule of assertion is much like the constitutive rules of chess, in that, the rules of chess constitute the game of chess itself. That is, it is not possible to play chess unless the rules of chess are in force and being observed. Accordingly, assertion is constituted purely by the rule that governs it. Furthermore, he claims that the rule of assertion is much like the constitutive rules of chess, in that, such constitutive rules do not lay down necessary conditions for the performance of an act. That is, the violation of a rule does not result in ceasing to play the game. Accordingly, if a speaker tells a lie using an assertion, and thus wittingly violates (KA), they nevertheless perform an assertion. In an attempt to clarify the notion of a constitutive rule (or norm) for assertion, Williamson writes:

> Constitutive rules are not conventions. If it is a convention that one must Φ, then it is contingent that one must Φ; conventions are arbitrary and can be replaced by alternative conventions. In contrast, if it is a constitutive rule that one must Φ, then it is necessary that one must Φ. More precisely, a rule will count as constitutive of an act only if it is essential to that act: necessarily, the rule governs every performance of the act. (490)

On this view, then, what makes an assertion the type of speech act that it is, rather than some other type of speech act, is that the speaker says that *p* and their saying is governed necessarily by a constitutive rule (or norm), the condition of which is knowledge.[12]

---

[12] Knowledge is arguably the most popular condition for proponents of the constitutive rule (or norm) account. However, as Whiting (2013) observes, there is now something of a cottage industry for generating conditions for

2.2. **Assertion in the Epistemology of Testimony**

In epistemology, the speech act of assertion is treated as an instance of 'testimony', which itself has become a philosophical term of art, intended to capture a diverse array of cases in which an individual makes an utterance which may serve as a source of knowledge for another individual. For example, the term may extend over cases of every-day conversation, formal courtroom testimony, secret diary entries, overheard soliloquy, classroom lectures, examination answers, social-media posts, etc. As such, there are cases of testimony that are oral or written, individual or institutional, private or public, anonymous or attributable, etc. Due to the diverse array of cases 'testimony' is intended to capture, especially private cases (e.g., secret diary entries and overheard soliloquy), where the speaker is typically not understood as directing their thoughts and feelings towards another person, some have questioned whether testimony is a unitary category. As Lackey (2008) observes:

> On the one hand, we often think of testimony as a source of belief or knowledge for hearers, regardless of the speaker's intention to be such a source. On the other hand, we often think of testimony as involving the intention to communicate information to other people, regardless of the needs or interests of the hearers. (19)

As such, Lackey thinks we should distinguish between what she calls 'speaker testimony' (or '*s*-testimony') and 'hearer testimony' (or '*h*-testimony'):

> *S*-testimony: *S* gives speaker testimony that *p* if and only if, by performing an act of communication *a*, *S* reasonably intends to convey the information that *p* (in part) in virtue of *a*'s communicable content.

---

the constitutive rule (or norm) of assertion, e.g., truth, belief, justification, justified belief, true justified belief, reasonableness to believe, and so on.

> *H*-testimony: *S* gives hearer testimony that *p* if and only if, by performing an act of communication *a*, *S*'s audience reasonably takes a as conveying that information that *p* (in part) in virtue of *a*'s communicable content. (Gelfert 2014: 35, which paraphrases and simplifies Lackey 2008: 30–32)[13]

To clarify, for Lackey, both cases refer to testimony given by the same speaker *S*, who is performing one and the same act of communication *a* (e.g., penning a diary entry or expressing propositions aloud solitarily). The difference, as she sees it, lies in the perspectives of the speaker and the hearer, specifically between the speaker's intention to communicate and the audience's comprehending an act of communication as conveying content. The idea, then, is that an individual who keeps a secret diary, for example, and so seemingly has no intention of conveying the information to another person, while not giving *s*-testimony, may nonetheless be understood as giving *h*-testimony, if some reader reasonably takes the diarist's writings as conveying information in virtue of their communicable content. According to Lackey, since a speaker's intention and the hearer's interpretation of an act of communication can plausibly come apart in this way, we should opt for a disjunctive view of the nature of testimony, as follows:

---

[13] Lackey (2008: 28–29) uses 'communication' in a very broad way. That is, for her, communication does not necessarily involve an interpersonal relation between a speaker and a hearer, but merely requires that 'the speaker intend to express communicable content'. Indeed, it might be thought that it would be better for her to define testimony explicitly in terms of the broader notion of an intention to express communicable content, rather than communication. That is not to say, though, that her notion of intending to express communicable content is itself entirely clear, but only to say that communication may not be the best notion to employ in a definition of testimony which is designed to include cases in which there is no interpersonal relation between a speaker and a hearer. Note further that she provides three instances of what it is for an act of communication, *a*, to 'convey the information that *p*': (i) where *a* is the utterance of a declarative sentence such that it expresses the proposition that *p*, *a* conveys the information that *p*; (ii) where *p* is an obvious (uncancelled) pragmatic implication of *a*, *a* conveys the information that *p*; and (iii) where an act of communication *a* expresses the proposition that *q*, and it is obvious (either to everyone in the exchange or to a normal competent speaker) that *q* entails *p*, *a* conveys the information both that *q* and that *p*.

> The disjunctive view of testimony: *S* testifies that *p* by making an act of communication *a* if and only if (in part) in virtue of *a*'s communicable content, (1) *S* reasonably intends to convey the information that *p* or (2) *a* is reasonably taken as conveying the information that *p*. (35–36)

The epistemology of testimony, then, is concerned with the way in which testimony facilitates the epistemic relationship between an individual and a fact (or true proposition). Specifically, it is concerned with the following two questions:

- How does testimony put an individual in a position to know the truth of what is testified?
- Under what conditions does testimony put an individual in a position to know the truth of what is testified?[14]

Although the speech act of assertion is treated as an instance of testimony, there are two issues that arise with respect to its status in the epistemology of testimony literature. First, it is not always clear which (if any) of the four broad accounts of assertion is being employed at any one time by epistemologists of testimony.[15] This may be because they wish to use the term 'assertion' in such a way which serves their specific purposes, and/or because they do not want to entangle themselves in the commitments of any particular account of assertion

---

[14] Wright (2018: 3–4) helpfully articulates the central concerns of the epistemology of testimony in terms of these two specific questions. To draw out the difference between them, he gives the example of iron rusting. As he says, an account of how iron rusts might appeal to the process of losing oxygen atoms, or the oxygen combining with the iron at an atomic level, whereas an account of under what conditions iron rusts might appeal to the presence of oxygen, iron and moisture. Although the two questions are not independent of each other – since knowing the answer to one might enable someone to infer the answer to the other – they are nevertheless conceptually different.

[15] See, for example, Gelfert (2014) and Hinchman (2020). For exceptions to this, see Fricker (2015) and Goldberg (2015), who endorse the (or a version of the) constitutive rule (or norm) account of assertion explicitly in their work on testimony.

from the philosophy of language.[16] Second, and related to the first, it is not always clear what the relation is between testimony and assertion for epistemologists of testimony.[17] As Goldberg (2010) observes, some appear to identify testimony with assertion: a speech act constitutes testimony if and only if it is the speech act of assertion.[18] Some, on the other hand, more explicitly endorse what he calls the 'necessity thesis': a speech act constitutes testimony only if it is the speech act of assertion.[19] Whether epistemologists of testimony opt for the identity relation or the necessity thesis, it puts pressure on which (if any) of the four broad accounts of assertion from the philosophy of language can be utilised in the epistemology of testimony, for 'assertion' would seemingly need a considerably broad extension in order to capture the diverse array of cases of testimony in the epistemology of testimony literature. Goldberg (2015), however, argues that the necessity thesis, and thus the identity relation, do not hold. He starts with the intuitive claim that the speech act of speculation is distinct from the speech act of assertion. He then demonstrates that it is nonetheless possible to treat the speech act of speculation as a case of testimony, using the following example:

When Harry Met Sally

> A hearer, Harry, knows that a given speaker, Sally, has extremely high epistemic standards. In particular, Harry knows that Sally asserts something only if she is absolutely (epistemically) certain. (Sally rarely asserts anything, but when she does

---

[16] For example, they may use 'assertion' to refer to only paradigmatic cases of the speech act. (See §2.5 for more on the paradigm case of assertion.)

[17] Goldberg (2010; 2015) is one exception since he has attempted more than most to clarify the relation between testimony and assertion using his version of the constitutive rule (or norm) account of assertion. Hinchman (2020) also provides a unique approach to the relation between assertion and testimony, though I do not have space to address it here.

[18] See, for example, Fricker (1987) and Sosa (1994). Note, Goldberg says 'appear to' because their comments are made in passing, and so the attribution is not certain.

[19] See Hinchman (2005), Moran (2005; 2018) and Owens (2006),

> you can bet your house that it is true.) Harry also knows that Sally's high epistemic standards inform her speculations as well: Sally will not speculate that *p* unless [*p*] is true. (Harry knows that even when speculating, Sally hates to be seen as ever having presented-as-true something that is false.) In fact, Harry knows that Sally's track record vis-à-vis her speculations is as good as the track record of highly competent asserters. So when Harry observes Sally speculate that *p*, under conditions in which Harry has all of this background knowledge regarding Sally's speech dispositions, Harry thereby comes to believe that *p* on the basis of Sally's having so speculated. Unsurprisingly, [*p*] is true. What is more, and equally unsurprisingly, Sally's speculation was in fact both sensitive – Sally would not have so speculated had [*p*] been false – and reliably formed – a preponderance of the speculations she would arrive at through this same process would be true. Finally, there are no relevant defeaters. (55)[20]

According to Goldberg, under these conditions it seems obvious that we can credit Harry with coming to know that *p* on the basis, or at least partly on the basis, of accepting Sally's speculation. As he says, we should not deny this merely because Sally's contribution is a speculation, since Harry already knows this, and has taken it into account. Given, then, that it is seemingly possible to treat the speech act of speculation as a case of testimony, the necessity principle, and thus the identity relation, apparently do not hold. However, this of course does not explain away the issue that it is not always clear what the relation is between testimony and assertion for epistemologists of testimony. Indeed, it only underscores it.

2.3. **The No-assertion View**

As we have seen, both philosophers of language and epistemologists of testimony work on the assumption that the term 'assertion' picks out an act type that we actually engage in, and

[20] For a helpful exposition of the nature of epistemic defeasibility and defeaters, see Pollock (1987).

which delineates a category we need in order to explain a significant component of human behaviour. However, this assumption has recently been called into question by proponents of what is known as the 'no-assertion view'. According to Cappelen (2011; 2020a), the term 'assertion' not only fails to pick out an act type that we actually engage in, but it does not delineate a category we need in order to explain any significant component of human behaviour. Indeed, he thinks the concept of assertion is defective and thus that the entire topic of assertion is illusory and should be ignored. While this view might seem radical, bold, and provocative, he maintains that it is not. Why? Because, he argues, almost every theoretical discipline will, at core junctures, leave behind some ordinary language classifications. He writes:

> Theoretical physics […] has little use for 'thing' or 'pizza' or 'duck'. Contemporary syntactic theory has little use for 'subject' and 'predicate'. In the theory of meaning, the ordinary word 'meaning' should be left behind and replaced with more refined notions. It doesn't have to be like that, but if it happens, as disciplines progress, bits of ordinary language fall out of the way we talk about that discipline. It's not radical or rare. On the contrary, it happens all the time when theorizing gets more advanced. The No-Assertion view claims that 'assertion' is like 'duck' to physics or 'predicate' to syntax. It is an unimportant category when trying to understand language, speech and communication. (2020a: 139)

Cappelen sees the elimination of 'assertion' as less troublesome than the analogous cases above since the term plays only a marginal role in common sense descriptions of human behaviour. That said, he concedes that eliminativism about assertion is radical from the point of view of contemporary theory, given that it goes against an entrenched tradition in the philosophy of language over the last 110 years or so. Nonetheless, he submits that all the work done by appeals to assertion in that tradition is better done without such appeals. The core components of Cappelen's no-assertion view are as follows:

- There are sayings and these are the sorts of things Austin described as locutionary acts.

- Sayings are evaluated by contextually variable norms, they have variable causes and effects, and they are accompanied by contextually variable commitments. None of these are constitutive of the speech act of saying. (140–141)

Notice, then, that 'saying' is a shared notion not only among proponents of the pro-assertion view but among proponents of the no-assertion view as well. Cappelen, however, sees proponents of the pro-assertion view as proponents of different versions of what he calls the 'saying+ view' of assertion, i.e., the general view that assertion involves the locutionary act of saying plus some other feature which is supposedly essential to it.[21] According to him, saying+ views always proceed in three stages. First, a subset of sayings is identified in terms of certain causes (i.e., the attitudinal account), certain effects (i.e., the common ground account), certain commitments (i.e., the commitment account), or certain rules (or norms) (i.e., the constitutive rule (or norm) account). Second, having carved out an act type in terms of saying plus whatever feature, the added feature is deemed an essential element of that act type, in that, it could not be that act type if it did not have this particular feature, whether it be a certain cause, certain effect, certain commitment, or certain rule (or norm). The label 'assertion' is given to this act type. Third, the act type of assertion is then put to work in order to play some apparently important role in theorising about speech, communication, and the spread of knowledge. Proponents of the no-assertion view are opposed to the saying+ view and thus this three-step process. Why? Cappelen (144–146) provides four arguments.

2.3.1. **The Simplicity Argument**

Cappelen maintains that there are two dimensions along which the no-assertion view is simpler than the saying+ view. First, he argues that the saying+ view involves playing the 'assertion counterexample game', whereas the no-assertion view does not. The assertion counterexample game is where a proponent of the saying+ view, having identified a subset

[21] From now on, I use 'saying+ view' to refer to all pro-assertion views.

of sayings in terms of certain causes, certain effects, certain commitments, or certain rules (or norms), and essentialising and labelling the resulting act type 'assertion', is faced with a counterexample designed to show that it is possible to make an assertion which does not involve that essential feature. The proponent of the saying+ view must then explain away the counterexample, to safeguard their saying+ view. The no-assertion view, on the other hand, predicts for any '+' proposed that there will be a saying which does not confirm to this particular '+', because any causes, effects, commitments, or rules (or norms) are only contingently related to sayings. Second, Cappelen argues that the saying+ view, unlike the no-assertion view, is more complicated, because it makes claims about what happens in all possible worlds, i.e., it makes a modal claim. For example, Williamson's constitutive rule (or norm) account, as we saw in §2.1.4, holds that the speech act of assertion is necessarily governed by (KA). According to Cappelen, Williamson needs 'modal evidence' to establish what happens in every possible world, modal evidence which is apparently hard to get and is more or less totally lacking in Williamson's account. In other words, the no-assertion-view is simpler because it does not make modal claims and thus is not burdened with having to provide modal evidence. He concludes that the simplicity of the no-assertion view compared to the saying+ view makes the no-assertion view more appealing.

2.3.2. **The Variability Argument**

Cappelen maintains that there is variability in the causes, effects, commitments, and rules (or norms) of sayings, which undermines the saying+ view. For example, he claims, in the relevant contexts, it is fine for speakers to say things just to be funny or entertaining, or to say things they do not believe or have good evidence for, or to say things they are not committed to defending when challenged, or to say things they do not want added to the common ground. In other words, he argues that sayings take place in settings where there are variable causes, variable effects, variable commitments, and variable rules (or norms), but where no one of these plays a particularly important role over the other, and so any attempt to pick out any one of these subsets as capturing the essence of assertion is pointless. He concludes that it is not possible to formulate a meaningful and unified category under the label 'assertion'.

2.3.3. **The Explanatory Power Argument**

Cappelen maintains that the no-assertion view can explain all the data that the saying+ view tries to explain but without making the same commitments as the saying+ view. He accepts that subsets of sayings exist which involve certain causes, certain effects, certain commitments, or certain rules (or norms). However, he argues that since none of these features are essential to any particular act on the no-assertion view, the modal component burdening the saying+ view is removed. Moreover, he argues that the modal claim at the heart of the saying+ view does not do any explanatory work anyway, but only creates complications, which are brought out by the assertion counterexample game. He concludes that the saying+ view is less appealing than the no-assertion view because it burdens itself with modal commitments without gaining any explanatory power over the no-assertion view.

2.3.4. **The Argument from the Method of Elimination**

Drawing on Chalmers's (2011) work on verbal disputes, Cappelen maintains that the term 'assertion' can be eliminated legitimately from philosophical theorising without issue. He observes that the literature on assertion contains different accounts of assertion but claims that there is a distinct possibility that participants in the debate are talking past each other, i.e., they are not talking about the same thing and do not have a substantive dispute. He then suggests, using Chalmers's procedure for eliminating verbal disputes, that we can test this by eliminating the use of the term 'assertion' and describing the disagreement in assertion-neutral terms. He predicts that the disagreement can be described purely by appeal to the notion of sayings and their various causes, various effects, various commitments, and various rules (or norms). He concludes that this is the level at which the debate should be conducted, that it would be conducted with very little substantive disagreement, and that what remains would be easily resolved.

Cappelen acknowledges that the no-assertion view is destructive and critical in as much as its objective is to eliminate the category of assertion from the theory of speech acts.

However, he believes that it can be looked at in another, more positive light. Specifically, he sees it as leading to what he calls a 'saying-first view' of speech act theory. Speech act theory, he claims, has been dominated by efforts to develop various classifications of illocutionary acts on the presupposition that there are tidy and useful classifications of locutionary acts. The saying-first view rejects this presupposition. This, he envisions, opens up an alternative research field with three focal points:

- The nature of sayings.
- The complex ways in which the various causes, various effects, various commitments, and various rules (or norms) of sayings interact in particular cases.
- The evolution of sayings over time, the difference across cultures and contexts, and the intentional revision of the various causes, various effects, various commitments, and various rules (or norms) of sayings. (Cappelen 2020a: 153)

In sum, if the no-assertion view is correct, then the whole enterprise of trying to articulate the nature of the speech act of assertion is entirely misguided, and a significant revision of speech act theory is potentially required.

2.4. **Conceptual Engineering: Carnapian explication**

Ultimately, I think the no-assertion view fails. However, I do think it sheds light on an uncomfortable truth: the assertion debate has lost its way. That is, it seems only fair to go some distance with Cappelen and admit that what 'assertion' denotes in the philosophy literature generally is not entirely clear and thus its philosophical application uncertain. Why has this happened? Mainly, it seems, because the literature on assertion within the philosophy of language developed from interests in a variety of different language-based issues, by means of a variety of different apparent cases of the speech act. And, in addition to that, but largely detached from it, the literature on assertion within the epistemology of testimony developed from interests in a variety of different epistemological issues, again, by means of a variety of different apparent cases of the speech act. As such, philosophers now find themselves asking, 'What is assertion?', as though it is a difficult question about the

underlying nature of some determinate philosophical category. So, where do we go from here? It seems we have three options:

(i) Submit to the no-assertion view.
(ii) Push for a 'family resemblance view' of assertion, according to which 'assertion' denotes a spectrum of speech acts all connected by a series of overlapping similarities, rather than some essential feature(s).[22]
(iii) Endorse the saying+ view but recognise that philosophy simply has not made up its mind about how the concept of assertion is best understood, and thus realise that what must be sought is a proposal for how it might fruitfully be applied.

I opt for (iii). Why? Two reasons. First, I think the arguments for (i) are unpersuasive, as I discuss in §2.6. Second, against (ii), I hold the general position that a systematised theory of speech acts is desirable, because a clear-cut taxonomy of speech acts would be useful in the philosophy of language and epistemology, and possibly further afield as well. However, as a proponent of the saying+ view who recognises that philosophy has not made up its mind about how the concept of assertion is best understood, and who believes that what must be sought is a proposal for how it might fruitfully be applied, I think a new approach is needed. Specifically, I think the concept of assertion is ripe for 'conceptual engineering'.

'Conceptual engineering' is a relatively new term which refers to a set of activities the general purpose of which is to assess and ameliorate our concepts, our core representational devices.[23] The brand of conceptual engineering I think most appropriate for assertion is Carnap's (1947; 1950) 'explication', which is a well-established and independently plausible account of concept assessment and amelioration. It is a model of how philosophers *should*

[22] For more on family resemblance concepts, see Wittgenstein (1953).

[23] See e.g., Clark & Chalmers (1998), Haslanger (2000), van Inwagen (2008), Plunkett & Sundell (2013), Scharp (2013), Leslie (2017), Simion (2017; 2019), Burgess, Cappelen & Plunkett (2020), Cappelen (2020b), and Thomasson (2020).

engineer concepts, rather than a description of how they *in fact* engineer concepts. Specifically, Carnapian explication is the replacement of one or more vague or indeterminate concepts, the 'explicandum' or 'explicanda', with more exact counterparts, the 'explicatum' or 'explicata'. Explicanda can be concepts from an earlier stage of theorising or folk concepts, and explicata can (but need not) be expressed using the same words as corresponding explicanda. Carnap (1950) stipulates that an explicatum is to satisfy four requirements to a sufficient degree:

- The explicatum is to be *similar* to the explicandum 'in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used'. (7)
- The explicatum is to be *precise*, or at least more precise than the explicandum, in virtue of explicit rules of use (or definition) given in terms of a 'well-connected system' of concepts. (ibid)
- The explicatum is to be *fruitful*, in the sense that it facilitates the 'formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept). (ibid)
- The explicatum is to be *simple*.

Here is an example (from Pinder (2020a)). Last century, there was no accepted definition for the concept of a planet. It was explicitly accepted, though, that the concept PLANET had nine canonical instances, Mercury through to Plato, and a number of canonical non-instances, such as the Sun, the Moon, asteroids, etc.[24] However, in the early twenty first century, astronomers discovered several 'trans-Neptunian objects', such as Eris. These objects were similar in relevant respects both to Pluto and Eris, and thus seemed to be borderline cases of PLANET. As a result, in 2006, the *International Astronomical Union* (IAU) engineered a new concept, explicitly defined in terms of key properties of celestial bodies, to replace PLANET and facilitate a more principled classification of Pluto and other celestial

[24] I use capitals to denote the concept.

objects. The IAU considered two candidate explicata, PLANET* and PLANET**, defined roughly along the following lines:

PLANET*

A celestial object *x* falls under PLANET* iff:

(i) *x* orbits the Sun
(ii) *x* is sufficiently large for its own gravity to have formed it into a sphere.

PLANET**

A celestial object x falls under PLANET** iff:

(i) *x* orbits the Sun
(ii) *x* is sufficiently large for its own gravity to have formed it into a sphere
(iii) *x* has cleared its neighbourhood of debris.[25]

Following a vote, the IAU replaced PLANET with PLANET**, since both Pluto and Eris fall under PLANET*, but neither fall under PLANET**. Its members now use PLANET** for celestial taxonimisation, instead of using PLANET. As such, the IAU count as having explicated PLANET.[26] We might ask the following question, then: did the IAU make the

[25] See International Astronomical Union (2006a; 2006b).

[26] Note, it is a mistake to think that explication requires us to introduce formal systems for concepts. For example, Strawson (1963: 503) writes that '[the] method is to construct a formal system, which uses, generally, the ordinary apparatus of modern logic and in which the concepts forming the subject-matter of the system are introduced by means of axioms and definitions'. However, Carnap (1947: 8) explains that explication involves 'replacing [a vague or not quite exact concept] by a newly constructed, more exact concept'. As Pinder observes, one advantage of understanding explication without this requirement is that it better reflects actual scientific and philosophical practice, which often involves the construction of concepts that are informal or, at least, not fully

right decision in replacing PLANET with PLANET**? The answer to this question of course depends on whether PLANET** satisfies Carnap's four requirements to a sufficient degree. For a deeper understanding of the method of explication, it is worth taking a moment to answer this question. However, to do that, some further comments on Carnap's four requirements are in order.

Typically, it is thought that Carnap prioritised the fruitfulness requirement over similarity, precision, and simplicity.[27] Notice, though, that there are at least two ways to interpret this thought.[28] First, fruitfulness might be weighted more heavily than the other desiderata, so that a small increase in fruitfulness is preferable to a larger increase in one of the other desiderata. Second, fruitfulness might subsume other desiderata so that precision and simplicity in particular are only desirable as a means to fruitfulness. For our purposes, it does not matter which interpretation we choose, for whichever way we interpret this thought, fruitfulness is the principal aim when undertaking an explication. That said, as Kitcher (2008) observes, Carnap's conception of fruitfulness is problematic, in that it is too narrow: if fruitfulness is solely understood in terms of facilitating the formulation of universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept), not only must similarity take priority over fruitfulness, but fruitfulness is inappropriate for many areas of theoretical inquiry.[29] Pinder (2020a) attempts to remedy this problem, however, by tying fruitfulness to the specific 'theoretical goals' that led to the introduction of the explicatum. I am sympathetic to Pinder's position, so it is worth fleshing it out in some detail. He calls his view the 'relevant-goals account of fruitfulness', the tenets of which are as follows:

---

formalised. Indeed, the explication of PLANET is a good example of this. For more on this, see Brun (2016) and Dutilh Novaes & Reck (2017).

[27]See Schupbach (2017) and Dutilh Novaes & Reck (2017).

[28] See Kitcher (2008) and Pinder (2020a).

[29] As Kitcher (2008: 115) writes, Carnap's proposal 'doesn't suit parts of the physical sciences, and [is] deeply problematic for the biological, earth, and human sciences'. Such sciences do not trade in the kind of universal statements that are more familiar in, say, classical physics and mathematics.

- An explicatum is fruitful insofar as its replacement of the corresponding explicandum would facilitate, through the ordinary course of inquiry, progress towards achieving relevant theoretical goals.
- Given multiple fruitful candidate explicata for a single explicatum, the most fruitful candidate is that whose replacement of the explicandum would facilitate, through the ordinary course of inquiry, most progress towards achieving relevant theoretical goals. (6)

According to Pinder, there are two principal ways that progress should be understood as facilitated towards achieving theoretical goals.[30] First, progress is facilitated if the goal is achieved by the very act of replacing the explicandum with the explicatum. Second, progress is facilitated if a step is made towards achieving the goal by the act of replacing the explicandum with the explicatum.[31] As such, when deciding between multiple fruitful candidate explicata, we should compare how much progress would result from replacing the explicandum with each candidate. Pinder claims that there is no general algorithm for such comparisons and relative fruitfulness must be assessed, using academic judgement, on a case-by-case basis. It follows from this that different theorists, even those who share theoretical goals, may make different judgements on a given case. However, it does not follow from this that any one of the theorists is mistaken necessarily. There is perhaps a subjective element to such judgements, after all. What, then, are theoretical goals? Focusing on scientific theoretical goals specifically, i.e., goals the achievement of which would lead to progress in scientific inquiry, Pinder provides the following examples:

---

[30] Of course, what constitutes progress in inquiry plausibly varies between different fields.

[31] Note, Pinder holds that there is no requirement that those performing the explication know that such a step has been successfully made so long as, in typical cases, theorists have a good idea of what counts as a move in the right direction. He notes further that to say that progress is facilitated through the 'ordinary course of inquiry' is to rule out atypical cases in which an explication results in progress towards achieving relevant theoretical goals in deviant cases. For example, there may be cases in which a 'bad' explication nonetheless inspires a theorist to look at a problem in a new and illuminating way.

- To explain some phenomenon
- To formulate universal generalisations
- To formulate empirical laws
- To develop a taxonomy
- To provide a principled extension for a term. (7)

As he points out, a notable feature of such goals is that, if achieved, they might genuinely lead to better scientific theories. That is, if phenomenon *P* falls within the scope of some scientific theory, then that theory is ceteris paribus better if it explains *P*. All else being equal, a scientific theory that includes (true) universal generalisations or empirical laws is better than one that does not. Thus, he recommends that we understand theoretical goals in terms of 'theoretical values'. He writes:

> Theoretical values [for scientific inquiry] are general characteristics of good scientific theories, such as: internal consistency; coherence (with other accepted theories); evidential accuracy (i.e., fit with the evidence); scope (i.e., applicability to a wide range of phenomena and/or cases); explanatory power; simplicity; and so on. Roughly, the more theoretical values a theory has, and the greater extent to which it has them, the better the theory. (8)[32]

Accordingly, Pinder gives an account of theoretical goals in terms of theoretical values, which he calls 'value directedness':

[32] Note, even Pinder admits that there are complications here, since some values might be weighted more heavily than others (such as internal consistency over simplicity), and values might sometimes conflict (increasing scope can decrease evidential accuracy). Moreover, there are also important questions about what the theoretical values in fact are, how they are related, and why they are valued. For present purposes, however, like him, I put such complications and questions aside.

> A goal is theoretical with respect to *T* [where *T* is a scientific theory] insofar as: ceteris paribus, achievement of that goal would ipso facto transform *T* into a theory that is better with respect to some theoretical value(s) than *T*. (9)

There is one obvious question that arises from all this: which theoretical goals are relevant? Pinder proposes a contextualist response to this question. He reasons as follows. Explications are performed for particular theoretical purposes and fruitfulness is the principal desideratum for explicata. As such, it makes good sense to measure fruitfulness against those purposes. However, for any given explication, the relevant theoretical goals are the explicator's theoretical goals, where the 'explicator' may be a theorist or an institution. If we suppose that an individual theorist is performing an explication, then explicator's theoretical goals are the personal theoretical goals of that individual theorist *qua* performer of the explication. If we suppose that an institution is performing the explication, then the explicator's theoretical goals are the institutional theoretical goals of that institution *qua* performer of the explication. Pinder concedes that, on his contextualist picture, explicators have a significant degree of control over which of their theoretical goals are relevant, and thus over the measure of fruitfulness that applies in a given case. However, he maintains that this does not imply that anything goes. Why? He provides two reasons. First, he claims, relevant theoretical goals must (by value-directedness) be directed towards the improvement of certain theories in certain respects – an explicator cannot simply decide to uniformly degrade theories in the name of fruitfulness. Second, he claims, additional practical, personal, normative, etc. factors, external to the method of explication, constrain an explicator's theoretical goals.

With all this set out, Pinder demonstrates the utility of his relevant-goals account of fruitfulness by answering the question above: did the IAU make the right decision in replacing PLANET with PLANET**? According to him, the IAU did make the right decision. Let us see why. As mentioned above, the IAU considered two candidate explicata when explicating PLANET: PLANET* and PLANET**. Using its website as source, Pinder extracts the IAU's institutional theoretical goals in explicating PLANET:

a) To provide clearly defined astronomical nomenclature.
b) To provide a taxonomy for celestial objects that reflects our current understanding (with specific reference to 'planet').[33]

As such, in order to determine which candidate promised to be most fruitful, we must consider the extent to which replacing PLANET with PLANET* or PLANET** facilitates progress towards achieving (a) and (b). Let us consider (a). PLANET does not serve to underpin clearly defined nomenclature. Why? Since the concept is defined only by heterogeneous instances and non-instances, it fails to classify many celestial objects in our solar system and provides no principled basis for extension to objects in other solar systems. In contrast, in virtue of their explicit definitions, both PLANET* and PLANET** are well-placed both to classify the celestial objects in our solar system and extend to objects in other solar systems. Thus, by replacing PLANET with either candidate explicatum, the IAU would take a clear step towards achieving (a).[34] Let us consider (b). In light of the discovery of Eris and other trans-Neptunian objects very similar to Pluto, PLANET does not reflect contemporary understanding of the solar system. In contrast, both PLANET* and PLANET** are defined in terms of well-understood celestial properties and so replacing PLANET with either candidate explicatum constitutes a step towards achieving (b). So, we can see that PLANET* and PLANET** satisfy Carnap's four requirements for explication. First, PLANET* and PLANET** are sufficiently *similar* to PLANET. Second, PLANET* and PLANET** are more *precise* than PLANET. Third, PLANET* and PLANET** are *fruitful*. Fourth, PLANET* and PLANET** are *simple*. However, PLANET** is more fruitful than

---

[33] See International Astronomical Union (2006a; 2006b). Other aims include: to organise scientific meetings; to promote educational activities; and to discuss the possibility of 'future international large-scale facilities'.

[34] As Pinder notes, neither PLANET* nor PLANET** underpins a *perfectly* clear nomenclature. There are some difficulties spelling out, for example, how spherical an object must be in order to satisfy condition (ii) of both PLANET* and PLANET**, or to what extent an object must have cleared its object of debris to satisfy condition (iii) of PLANET**. Either of these difficulties could, in principle, have led to unclarity in nomenclature as more celestial objects were discovered, but there is no immediate reason to think that either candidate explicatum would be more likely than the other to lead to such problems.

PLANET*. Why? Because PLANET** contains an additional clause, namely, that a celestial object *x* must be such that *x* has cleared its neighbourhood of debris. While no celestial object in our solar system has cleared its neighbourhood of debris completely, it was known that there is significant difference between the extent to which Mercury through Neptune have cleared their neighbourhoods of debris, and the extent to which other celestial objects have cleared their neighbourhoods of debris.[35] As such, PLANET** reflects an important aspect of contemporary understanding of celestial objects that is not reflected by PLANET*. Replacing PLANET with PLANET**, rather than PLANET*, thus constitutes a greater step towards providing a taxonomy for celestial objects that reflect contemporary understanding. In other words, with respect to (b), PLANET** is more fruitful than PLANET*. Therefore, as Pinder claims, the IAU made the right decision in replacing PLANET with PLANET**.

Now, as mentioned above, I think Carnapian explication is the most appropriate form of conceptual engineering for assertion. Admittedly, however, the focus so far has been on scientific inquiry rather than philosophical inquiry. This, of course, raises a crucial question: does explication, especially Pinder's relevant-goals account of fruitfulness, generalise to non-scientific (e.g., philosophical) inquiry? As Pinder observes, in order to answer this question fully we would have to answer several other questions:

- What exactly are scientific theories?
- To what extent are they like the theories we find in other fields of inquiry?
- Are theoretical values applicable to non-scientific theories and, if so, to what extent?
- Does progress in other disciplines consist of developing better theories? (10)

That said, Pinder recognises that if we think of philosophy as broadly continuous with science, then we might think that the traditional theoretical values apply *mutatis mutandis* in philosophy. He adds a note of caution, however, that if we focus on, say, the normative nature of philosophical inquiry, then many of the traditional theoretical values may well be

[35] See e.g., Stern & Levison (2002) and Soter (2006).

inappropriate. For example, it is unclear that we can straightforwardly measure progress in ethics with reference to evidential accuracy or explanatory power. There are, then, deep issues about the nature of scientific inquiry and other inquiries, issues which I am unfortunately not in a position to resolve here. However, given that I use explication as a tool in providing a philosophical account of assertion, I am obliged to say something about its use in philosophical inquiry, or at least its use in my particular philosophical inquiry. I have three comments to make on this score. First, I do in fact think of philosophy as broadly continuous with science and so think that at least some of the traditional theoretical values of science apply *mutatis mutandis* in philosophy. Second, although my account of assertion does draw on some normative notions, it is not obvious that this is problematic from the outset, which leads to my next point. Third, I confess that I see my overall project as something of a trial run for the (de)merits of the application of explication in philosophical inquiry. It might be argued by some that it would be preferable to have an account of explication for philosophical inquiry set out in detail before undertaking the project I plan to undertake. However, we might think instead that the best account of explication is guided by what results we get if try to do something like we would in the scientific case. To rally some support on this, I call on Pinder, once more:

> [F]or the method of explication to be of practical value, it must simplify to some extent. A method that explicitly takes into account all of the complexities of inquiry would itself be too complex to provide helpful guidance. A simpler method, an idealisation that can be adapted on the fly to the situation at hand, is likely to be preferable. By analogy, speaking to a new driver in the UK, one might recommend that she drive on the left-hand side of the road. The driver can then adapt this advice as need be – for example, when navigating narrow lanes on which she is forced to drive in the middle of the road, when driving past parked cars which require her to move over to the right-hand side, when driving down multiple-lane one-way streets at the end of which she wants to turn right, and so on. It would not be helpful, so much as overwhelming, to try to take all of these complexities into account when offering the initial recommendation to the driver. (17)

With that, let us begin.

2.5. **Explicating ASSERTION**

Even though there is disagreement in the philosophy of language about what the speech act of assertion essentially is, and even though there is a lack of clarity in the epistemology of testimony about which (if any) of the four broad accounts of assertion is being employed at any one time and what the relation is between testimony and assertion, there is a point at which these two areas of philosophy intersect in a significant way on the topic of assertion. In the philosophy of language, there is a paradigmatic case of assertion. Similarly, in the epistemology of testimony, there is a paradigmatic case of testimony. Significantly, though, philosophers of language and epistemologists of testimony mutually agree that the paradigm case of assertion and the paradigm case of testimony are co-extensive. What, then, is a paradigm case of assertion? It might be thought that it is one in which a speaker says that *p* and intends that the hearer acquire a belief that the speaker believes that *p*.[36] However, as McDowell (1998) demonstrates, this cannot be correct. He asks us to consider the corresponding thesis about questions, namely, that the intention of an inquirer, in paradigm cases of inquiring, is to induce a hearer to induce in the inquirer a belief about the hearer's belief. Such a thesis, as McDowell points out, would be absurd. Why? Because the primary point of asking questions is not to find out what someone else believes but to find out how things are. As such, the primary point of making assertions is not for the speaker to instil in the hearer a belief about the speaker's own belief, but to inform the hearer about the subject matter of the speaker's assertion.[37] Here is an example of a paradigm case of assertion:

---

[36] See Williams (2002).

[37] Note, it is possible for '*p*' to be a placeholder for the proposition, 'I believe that *p*', anyway.

Bank Location

> Shannon is new in town and does not know where the bank is, so she asks a passing stranger, Sylvester, who informs her.
>
> Shannon: Excuse me, is the bank on the East- or West-side of town?
> Sylvester: It's on the East-side of town.
> Shannon: Thank you.

More formally, then:

> (AP) Paradigmatically, assertion is a communicative act which involves a speaker saying that *p* with the intention that the hearer acquire a belief that *p*, addressing the hearer's perceived need/desire for knowledge that *p*.[38]

Although the concept ASSERTION has one canonical or paradigmatic instance, philosophers of language and epistemologists of testimony respectively have introduced apparent borderline cases of assertion into the literature, cases which are similar in certain respects to the paradigm case but seemingly do not involve any intention on behalf of the speaker that a hearer acquire a belief.[39] To illustrate, here are three different types of apparent borderline cases of assertion:

---

[38] Unlike Lackey, I use 'communicative' (and 'communication') as we might typically expect, namely, as involving an interpersonal relation between a speaker and a hearer. In §4.2, I flesh this out using Grice's (1957; 1969; 1975; 1989) framework of communication.

[39] Again, I use capitals to denote the concept.

Cheating Student

> A Dean catches a student red-handed cheating during an exam. However, there is no direct confession from the student. The Dean has a strict policy by which she refuses to punish any student for cheating in the absence of a confession from that very student – which is well-known by all students. The Dean calls the student into her office and plays video footage of the incident. She then asks, 'Did you cheat?' The student says, 'No, I did not cheat'. The Dean sends the student away unpunished.[40]

Cases of this sort are known as 'bald-faced lies'. This type of case is described as one in which the speaker asserts that *p* even though it is common knowledge (between the speaker and the hearer) that the speaker does not believe that *p*. That is, the speaker and the hearer each know that the speaker does not believe that *p*, and each know that each know that the speaker does not believe that *p*, and each know that each know that each know that the speaker does not believe that *p*, and so on. It is this last point specifically, i.e., it being common knowledge (between the speaker and the hearer) that the speaker does not believe that *p*, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*. Here is a second type of case:

Sneaky Spy

> Suppose a spy, Natasha, asserts to her foreign counterpart that her partner is dead, yet she knows full well that he knows that her partner is alive and is being held by him in a nearby facility. Suppose, though, that the reason she asserts to him that her partner is dead is because she intends him to believe that she believes her partner is dead.[41]

---

[40] Adapted from Carson (2006).

[41] Adapted from Searle (1969).

Let us call cases of this sort 'double bluffs'. This type of case is described as one in which the speaker asserts that *p*, the speaker believes that not-*p*, and the speaker knows that the hearer knows that not-*p*. (Note, unlike cases of so-called 'bare-faced lies', it is not common knowledge (between the speaker and the hearer) that the speaker believes that not-*p*, since it is not the case that the speaker and the hearer know that the speaker knows that not-*p*.) It is this last point, i.e., that the speaker knows that the hearer knows that not-*p*, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*. Finally, here is a third type of case:

Interrogation Room

> In the course of their interrogation by the police, it must have become clear very quickly to the members of the Birmingham Six that nothing they could say or do would persuade their interlocutors either that they (the suspects) had not planted the bombs or that they (the suspects again) believed that they had not planted the bombs. For all that, when they uttered the words 'We did not plant the bombs', the suspects certainly asserted as much.[42]

Let us call cases of this sort 'resolute assertions'. This type of case is described as one in which the speaker asserts that *p* whilst at the same time knowing that the hearer will not believe what they assert. It is this last point, i.e., that the speaker knows that the hearer will not believe what they assert, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*.

Now, in order to conduct my explication, relevant theoretical goals must first be provided. My personal theoretical goals (as an individual theorist rather than an institution) are as follows:

---

[42] Adapted from Rumfitt (1995).

a) To provide a principled extension of the term 'assertion' for the philosophy of language and epistemology – indeed, for philosophy generally (and further field as well, if possible).
b) To facilitate the development of a taxonomy of philosophically interesting illocutionary acts for the broader theory of speech acts.[43]

As with scientific theories, a notable feature of such theoretical goals is that, if achieved, they might genuinely lead to better *philosophical* theories. That is, if phenomenon *P* falls within the scope of some philosophical theory, then that theory is ceteris paribus better if it explains *P*. As I see it, all else being equal, the philosophical theory that best balances *simplicity* (i.e., the least amount of complexity) with *strength* (i.e., the highest and most varied number of accurate predictions) is better than any other.[44] Thus, similar to scientific theories, we may understand theoretical goals in terms of theoretical values, where theoretical values for philosophical inquiry are general characteristics of good philosophical theories.[45]

---

[43] Two things. First, although these are my own personal theoretical goals, it seems they would apply to all proponents of the saying+ view. Second, although I have explicitly stated my personal theoretical goals from the outset, that is not to suggest that personal theoretical goals cannot be added, subtracted, replaced, or refined along the way. It would be artificial to think that explication does not have this dynamic nature. That said, even though I acknowledge this flexibility in explication, the reader should not expect a change in my personal theoretical goals.

[44] This thought is inspired by Lewis's (1973; 1980; 1994) 'best-system' theory for natural laws, according to which laws are the winners of a 'competition' among all collections of truths. Some truths are simple, e.g., the truth that grass is green. Some truths are strong, in that, they tell us a lot about the world. Typically, these are exclusive categories: simple truths are not strong, and strong truths are not simple. But there are some exceptions. For instance, the truth that any two objects are attracted to one another, with a force proportional to the product of their masses and inversely proportional to the distance between them, is relatively simple, but also quite strong in that it tells us a lot about the forces between many distinct objects. The laws, says Lewis, are these simple but strong truths.

[45] As Pinder notes, like scientific inquiries, this also includes internal consistency, coherence (with other accepted (or plausible) theories), evidential accuracy (i.e., fit with the evidence), scope (i.e., applicability to a wide range of phenomena and/or cases), explanatory power, simplicity, and so on. As we saw above (§2.4), evidential accuracy and explanatory power might not obviously apply to all philosophical theories. However, it is equally obvious

With this set out, what we need now is to select and put forward a candidate explicatum. My strategy, here, is to use (AP) to help us do this. Why this strategy? Because not only must any plausible explication of the concept ASSERTION classify the paradigm case, but the paradigm case is the point of least contention for gaining at least some understanding about what the term 'assertion' denotes in the philosophy literature. Moreover, by reflecting on the paradigm case, we can draw out the philosophically interesting phenomena related to it, phenomena which any plausible account of assertion must be able to account for. The significance of this last point will become clear shortly. Let us start, then, by reflecting on (AP) and making an initial list of features associated with it, as follows:

- Assertion involves the expression of a proposition, *p*.
- Assertion involves an interpersonal relation between a speaker and a hearer, whereby the speaker intends that the hearer acquire a belief that *p*.
- The general purpose of assertion is to inform the hearer (i.e., for the hearer to acquire knowledge that *p*).[46]
- Assertion is not essentially institutional (i.e., its performance does not depend on the existence of institutional positions or titles, e.g., like entering a plea in a court of law or sanctioning a marriage).[47]

What this initial list shows is that in order to provide an account of the paradigm case of assertion not only must we have one foot firmly in the philosophy of language, but we must have another foot firmly in epistemology, especially the epistemology of testimony. With that, let us build a more substantive list of the philosophically interesting phenomena

---

that the evidential accuracy and explanatory power apply to at least some (if not most) philosophical theories. As such, it is appropriate to include both here.

[46] Note, I say 'general purpose' because the speaker might of course be deliberately deceptive or simply unreliable. (Speaker (in)sincerity and speaker (un)reliability is discussed in §4.3.4.)

[47] Indeed, it seems that institutions and their practices are (at least in part) established by and sustained by means of paradigmatic instances of assertion. (This is discussed further in §5.3.)

related to these features, i.e., a list which pertains to its role in communication and the spread of knowledge:[48]

1) How assertion puts a hearer in a position to know the truth of what is asserted.
2) The conditions under which assertion puts a hearer in a position to know the truth of what is asserted.
3) Speaker sincerity and insincerity (i.e., the sense in which assertion relates to speakers' beliefs).
4) Speaker reliability and unreliability (i.e., the sense in which sincere assertion relates to the reliability of speakers' beliefs).
5) Expressing belief (i.e., the sense in which assertion expresses or manifests speakers' beliefs).
6) Representing oneself as knowledgeable (i.e., the sense in which speakers are typically thought to implicitly represent themselves has having knowledge in what they assert).
7) Commitment (i.e., the sense in which a speaker confers a commitment upon themselves to the truth of the proposition they assert).
8) Responsibility (i.e., the sense in which speakers incur a responsibility for the belief that the hearer is intended to acquire).
9) Challenges (i.e., the sense in which assertions can be contested by querying the speaker's epistemic standing on the proposition they have asserted).
10) Retractions (i.e., the sense in which a speaker can retract an assertion when they regard themselves as no longer in a position to believe or vindicate what they have asserted).
11) Moore's paradox (i.e., the apparent absurdity in attempting to assert '*p* but I do not believe that *p*').

[48] Note, Goldberg (2015) constructs a similar list when discussing what he thinks an account of assertion should be able to provide a vocabulary for. However, he does not explicitly construct his list in light of the paradigm case of assertion.

At this juncture, we should ask the following question: is this everything philosophers talk about when they talk about assertion? Answer: pretty much. Admittedly, there are other philosophically interesting phenomena associated with assertion, but compared to the phenomena on the substantive list they are peripheral in the study of the speech act. These include truth, semantics (i.e., the specification of meanings in linguistic expressions within a context), meta-semantics (i.e., the determination of semantic facts), implicature, presupposition, lying, so-called 'epistemic normativity', knowledge ascriptions, expert disagreement, testimonial injustice, radical interpretation, etc. Nonetheless, philosophers who work on assertion directly spend the majority of their time arguing about how assertion's core phenomena, i.e., the items on this substantive list, are best understood, while the list itself holds fixed.

Earlier, I said my strategy was to use (AP) to help us select and put forward a candidate explicatum. As mentioned above, I use this strategy not only because any plausible explication of the concept ASSERTION must classify the paradigm case, but because the paradigm case is the point of least contention for gaining at least some understanding about what the term 'assertion' denotes in the literature. Moreover, recall that I said by reflecting on the paradigm case, we can draw out the philosophically interesting phenomena related to it, phenomena which any plausible account of assertion must be able to account for. The reason why this last point is significant is because, as we have seen, almost all the philosophically interesting phenomena related to the speech act generally can be established solely in terms of the paradigm case. What I propose, then, is that our candidate explicatum for the concept ASSERTION should be nothing more than what the paradigm case denotes. In other words, we should shift from (AP) to (AE):

> (AE): Essentially, assertion is a communicative act which involves a speaker saying that *p* with the intention that the hearer acquire a belief that *p*.[49]

[49] Notice that in the shift from (AP) to (AE) I have dropped the following condition: 'addressing the hearer's perceived need/desire for knowledge that *p*'. This is because there are clear cases of assertion in which a speaker

I suspect some readers will find discomfort with my proposal given apparent borderline cases of assertion (like those above), cases (at least some of which) it might be felt are best understood as cases of assertion, but which do not fit (AE). Here, I would like to make a plea: since there is no accepted definition of the concept ASSERTION, but only an agreed paradigm case, let us not put our confidence in apparent borderline cases of assertion so hastily and wholeheartedly to guide our inquiry. As mentioned earlier, the literature on assertion within the philosophy of language developed from interests in a variety of different language-based issues, mainly by means of a variety of different apparent cases of the speech act. And, in addition to that, though largely detached from it, the literature on assertion within epistemology developed from interests in a variety of different epistemological issues, again, mainly by means of a variety of different apparent cases of the speech act. Due to this, the concept ASSERTION has become insipid (at best), if not defective (at worst). How, then, should we conduct our inquiry? Here, I would like to make a recommendation: our inquiry should be guided by, connected to, and framed within broader philosophical theories in order to provide a viable and principled account of the speech act.[50]

Let me explain this plea and recommendation. The idea that our candidate explicatum for the concept ASSERTION should be nothing more than what the paradigm case denotes is only a proposal – I do not claim to have provided a proof for it in any way. However, with this principled proposal in place, we can provide a theoretical elaboration of (AE), one which will afford us a vocabulary to explain almost all the philosophically interesting phenomena related to the speech act. Plus, there is no immediate reason to think that a theoretical elaboration of (AE) will not afford us a vocabulary to help us explain the outstanding peripheral phenomena as well. More than this, though, by providing a

---

asserts that *p* to a hearer who is not perceived by the speaker as needing/desiring to know that *p* (e.g., cases in which the hearer already knows that not-*p* and the speaker knows this but nonetheless asserts that *p* to the hearer to try and make the hearer lose or at least bring into question their knowledge that *p*). Thus, this condition does not capture an essential feature of assertion.

[50] See Keiser (2017) and Montminy (2020) for similar positions.

theoretical elaboration of (AE), we will be well placed to make principled verdicts on whether apparent cases of assertion really are best understood as cases of assertion or not. Of course, it is possible that (AE) will, in the end, show itself to be too restrictive and thus implausible. But, until a theoretical elaboration of (AE) is provided and the apparent cases of assertion scrutinised in light of it, we have no immediate reason to think this.

Notice that the preceding discussion draws out an important distinction at play in the methodology I employ here, namely, the distinction between a *candidate explicatum* and a *theoretical elaboration of a candidate explicatum*. This distinction is helpful because there can be agreement on a candidate explicatum but disagreement about how best to flesh out its finer details. With this in mind, consider again the four broad accounts of assertion above: the attitudinal account, the common ground account, the commitment account, and the constitutive rule (or norm) account. We might view these as theoretical elaborations of alternative candidate explicata for the concept ASSERTION. Notice, though, that one of them is not in competition with (AE), namely, the attitudinal account. That is, the idea that assertion essentially is when a speaker says that *p* and intends that their utterance be regarded by the hearer as a reason both to think that the speaker believes that *p* and that the speaker intends the hearer to believe that *p* also captures the paradigmatic instance of assertion and so is compatible with (AE). As I see it, this is a virtue, because it means the candidate explicatum I have suggested already has a history of independent plausibility. It might be thought, then, that my account of assertion should be labelled a defence of the attitudinal account. I have no serious qualms with such a label, but it is potentially misleading. Why? Two reasons. First, the theoretical elaboration of (AE) I develop throughout the subsequent chapters is different to any attitudinal account offered in the literature – at most, I put forward a *type* of attitudinal account, rather than a defence of any pre-existing attitudinal account. Second, my theoretical elaboration of (AE) in fact holds that all four broad accounts of assertion get something right, not just the attitudinal account. Specifically, the theoretical elaboration I develop for (AE) holds that assertion is best understood in terms of intentions, namely, an intention that the hearer acquire a belief (like the attitudinal account). But it is also consistent with the idea that the speech act counts as a proposal to the common ground (like the common ground account), involves the conferral

of a commitment on the speaker (like the commitment account), and has a significant and defining relationship with the epistemic position of the speaker (like the constitutive norm (or rule) account).

2.6. **A Defence of the Saying+ View**

What I aim to provide in this thesis is a novel theoretical elaboration of our candidate explicatum, (AE). Admittedly, by restricting the concept of assertion in this way, it will delineate a more modest sub-set of sayings and, in turn, a more modest set of contexts in which the speech act occurs than is sometimes thought. However, equipped with my theoretical elaboration, I hope to show that it can fruitfully be applied to almost all the philosophically interesting phenomena related to the speech act (i.e., all of items (1) through (11) on the substantive list in §2.5), as well as leave open the possibility that it may also fruitfully be applied to any outstanding philosophically interesting phenomena peripheral in the study of the speech act. This should not only vindicate the restriction of the concept, but it should also restore confidence in the concept itself.

Given that our inquiry will be guided by, connected to, and framed within broader philosophical theories, especially from the philosophy of language and the epistemology of testimony, I aim to be as transparent as possible with regard to the things that my theoretical elaboration of our candidate explicatum postulates, the nature of those things, and any other mechanisms involving them. Although our inquiry will be conducted in this way, that does not mean that apparent borderline cases from the philosophy literature will be entirely ignored. To be sure, proponents of the saying+ view must play the assertion counterexample game. However, if we prepare for the assertion counterexample game properly, i.e., by being equipped with a theoretical elaboration of our candidate explicatum, playing it is not the daunting prospect it is sometimes made out to be. In fact, it can be a useful philosophical tool. To reiterate, though, whereas some philosophers might hastily and wholeheartedly rely on arguments which turn on apparent borderline cases or apparent counterexamples as a way of guiding their inquiry, I am sceptical of this strategy. Instead, I think that arguments should turn on the theoretical elaboration of our candidate explicatum, which should

occasion philosophers to scrutinise apparent borderline cases or apparent counterexamples more cautiously. That is, with a theoretical elaboration of our candidate explicatum in hand, we should feel comfortable to raise questions about the inherent plausibility of apparent borderline cases or apparent counterexamples, or at least be willing to entertain the possibility that they might be best understood in terms of something other than assertion, especially if not doing so might all too quickly cost us the theoretical elaboration of the concept and the fruit it bears. I appreciate that some might worry that this will lead to an unpalatable revisionism about certain apparent cases of assertion. However, I think such a worry is misplaced. Why? Because such a worry makes sense only if the apparent cases of assertion under consideration are sufficiently well-established in the philosophy literature as cases of assertion in the first place in order to make revising them unpalatable. Yet, as we have seen, what 'assertion' denotes in the philosophy literature, apart from the paradigm case, is unclear.

Lastly, my theoretical elaboration of our candidate explicatum is rooted in observations about the sorts of creatures that we are, our aims, interests, and expectations (especially of each other), as well as the social practices we engage in to fulfil these. As such, any modal claims I make regarding the essential feature(s) of assertion are intimately tied to such observations. This, I hope, will mitigate any worries concerning any lack of modal evidence for any modal claims I make. Why? Because, as I see it, any modal claims I make regarding the essential feature(s) of assertion, given they are rooted in such observations, do establish what happens in all *relevant* possible worlds, namely, ones which are inhabited by the same or relevantly similar creatures as us, with the same or relevantly similar aims, interests and expectations (especially of each other) as us, and who engage in the same or relevantly similar social practices to fulfil these as us. Thus, the observations that reinforce the modal components I postulate, and by extension the modal components postulated, should show themselves to be rich in explanatory power.

In sum, with my explication, and indeed the theoretical elaboration which accompanies it, I hope to close the gap between the philosophy of language and the epistemology of testimony on the topic of assertion, offer up innovative solutions to problems pertaining to

assertion's role in communication and the spread of knowledge, and provide principled verdicts on whether certain apparent cases of assertion really are best understood as cases of assertion or not. In doing so, I hope to reaffirm the concept of assertion and, in turn, reinvigorate the assertion debate.

2.7. **Concession**

Although my overarching aim in this thesis is to explicate ASSERTION with (AE), it is of course possible that I will, in the end, be unsuccessful – such a replacement may never come to pass.[51] However, whether I am successful or not, one concession that the methodology I employ here affords us is, I hope, at least a plausible account of the paradigm case of assertion, which, in turn, I hope, will provide us with a unified set of plausible explanations for a litany of philosophically interesting phenomena pertaining to its role in communication and the spread of knowledge. In other words, even if my explication ultimately fails, my positions and arguments relating to this particular significant component of human behaviour will stand independently of that.

2.8. **Appendix A: outstanding questions for the conceptual engineer**

It would be disingenuous to proceed with our inquiry as if Carnapian explication itself does not have its own set of potential problems. As such, I deal with two objections typically levelled against Carnapian explication and conceptual engineering generally, to manage any immediate concerns.

[51] Interestingly, this might happen for legitimate reasons (e.g., not satisfying Caranap's requirements sufficiently, or not satisfying Carnap's requirements as well as some other candidate explicatum). Alternatively, it might happen for not so legitimate reasons (e.g., insufficient publicity or marketing).

2.8.1. **The Implementation Problem**

Suppose some term 'T' means M, but we attempt to conceptually engineer and in so doing attempt to revise the meaning of 'T', so that 'T' thereby means M*. The apparent issue is that such attempts are efforts to stipulate new semantic meanings and referents for terms whose semantic meanings and referents are already fixed when stipulation seemingly does not have this power.[52]

Deutsch (2020) is the leading proponent of this position. He claims, for example, that no single speaker, nor large groups of speakers acting together, can simply stipulate that, henceforth, 'dog' shall mean cats. Why? He gives two reasons. First, he argues that a speaker's stipulation that an existing term shall now have a different semantics does not suffice to give the term that new semantics. Second, he argues that everyone (including the conceptual engineers) is ignorant of what must be done, or what else must occur, in order to bring about such semantic changes, if it is even possible. Funnily enough, there is an air of stipulation in Deutsch's position, since he only poses a challenge or question to the conceptual engineer regarding the mechanism for implementation, rather than a serious problem for them.[53] In any case, I do not have a dog (or cat) in this fight, since Deutsch observes that there are certain cases in which a single speaker, or large groups of speakers acting together, can simply stipulate that some term 'T' means M* rather than M, namely, technical terms or terms of art. He writes:

> However, there is a class of important exceptions to the claim that the semantics of a term can't be simply stipulated as one pleases […] Sometimes we really do need new terms for things for which we lack conveniently short strings of letters or words. And not just in philosophy – in science, logic, mathematics, and everyday life, we

[52] For a comprehensive list of objections (and responses) to the method of conceptual engineering (including the two I discuss here), see Cappelen (2020b).

[53] For a response to this sort of challenge, see Pinder (2019).

> sometimes have use for convenient names, or predicates consisting of one or a few words, terms with which we can replace our clunkier descriptions […] There are countless cases like this in every theoretical context, cases in which something, or some property, important for theorizing, needs a convenient label. Stipulative introduction provides a straightforward way of introducing these labels and specifying their semantics. In other words, there are such things as "technical terms". Technical vocabulary often has its semantics fixed by stipulative introduction. Like all other theoretical endeavors, philosophy has its share of technical terms introduced via stipulative introduction. (3945–3946)

Given that 'assertion' is a philosophical term of art, we therefore need not worry about the power of stipulation in our effort to conceptually engineer it, for, by Deutsch's own admission, that power is granted to us purely by virtue of the kind of term that it is. That said, two comments are in order here. First, Deutsch claims that the stipulation of technical terms is not conceptual engineering proper. He writes:

> Somewhat surprisingly, it seems that some of what its defenders describe as "conceptual engineering" amounts to no more than the stipulative introduction of new terminology. I say "surprisingly" because there is nothing especially remarkable about stipulative introduction. It is certainly not some new, particularly fertile method of philosophizing, one that might supplement or supplant philosophical conceptual analysis. (3946)

I go some way with Deutsch on this point, in that, I think he is right to say that the stipulation of a technical term for the sake of philosophising is not remarkable. However, just because something is not remarkable does not mean that it is without value. When a technical term already exists in the literature which is vague in its reference and unclear in its application, either explication or elimination is a practical solution – 'assertion', as we have seen, is such a term. And if by stipulating a candidate explicatum we can regain some clarity in reference and application, and, in turn, reaffirm the concept and thus reinvigorate a philosophical debate, then our efforts will have been of value. Think of it as a tool for good

theoretical hygiene. Second, although 'assertion' is used as a term of art in philosophising, it is also used as an ordinary language term by the folk. This observation, however, is unproblematic. Why? For one, as Cappelen observes above, 'assertion' plays only a marginal role in common sense descriptions of human behaviour. For another, my proposal for what 'assertion' should denote, i.e., (AE), not only picks out paradigmatic instances of assertion in philosophical circles, but it also picks out paradigmatic instances of assertion in folk circles. I appreciate that some of the folk might have some intuitions about what 'assertion' refers to beyond paradigmatic instances, but it is a stretch to think that they will have any strong commitments either way. Moreover, they would likely have no compelling basis for such commitments. Relatedly, it seems reasonable to think that if my explication (or any other explication, for that matter) were to prove fruitful for the folk as well as philosophers, we could be more prescriptive about what 'assertion' denotes in folk circles based on the explication of the technical term within philosophy.

2.8.2. **The Topic Changing Problem**

Suppose some term 'T' means M, but to solve a philosophical problem we explicate and in so doing revise the meaning of 'T', so that 'T' thereby means M*. The apparent issue is that explicating 'T' does not result in solving the philosophical problem, but instead results merely in changing the topic.[54]

Strawson (1963) famously levels this objection against Carnapian explication. According to him, philosophical problems are solved by means of the philosophical clarification or illumination of concepts. The central philosophical problem he highlights is dealing with 'paradox and perplexity':

[54] Again, for a comprehensive list of objections (and responses) to the method of conceptual engineering (including the two I discuss here), see Cappelen (2020b).

> For it often happens that someone reflecting on a certain set of concepts finds himself driven to adopt views which seem […] paradoxical or unacceptably strange. (515)

Pinder (2020b) provides a helpful illustration, here. Suppose I take myself to be sat at home in my armchair. However, since some dreams are indistinguishable from sensory experience, my experience is compatible with my being in such a dream. Therefore, I do not know that I am sat at home in my armchair – I might only be dreaming that I am, after all. When we reflect on the concepts involved in this philosophical problem, especially KNOWLEDGE, the argument seems to be sound, but the conclusion seems to be unacceptable. Pinder calls this a 'conceptual imbalance'. For Strawson, then, we must do two things in order to solve such a conceptual imbalance. First, we must explain why a given step in the argument is in fact unsound, or why the conclusion is in fact acceptable. Second, we must explain why appearances are to the contrary. Thus, we solve the philosophical problem by resolving the corresponding conceptual imbalance. Enter Carnap, who writes:

> The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. (1947: 7–8)

The reason why Strawson takes issue with Carnap's approach is because he essentially sees it as turning your back on the philosophical problem – the old problems are stated in terms of the old concepts, after all. Strawson writes:

> [T]ypical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of use of exact and fruitful concepts in science. To do this last is not to solve the typical philosophical problem, but to change the subject. (1963: 506)

To be more precise, Strawson's objection is essentially this.[55] When we explicate a concept with an explicatum which is explicitly designed to be used for much narrower purposes, we *reconstruct* the conceptual imbalance. He concedes that this may result in the conceptual imbalance no longer arising: the argument no longer seems to be sound, or the conclusion no longer seems to be unacceptable. However, he argues that the explication will not tell us why the unsound step in the original argument seems to be sound, or else how or why the acceptable conclusion seems to be unacceptable, i.e., explication will not have told us why reflection on the concepts originally involved lead us astray in the first place. He concludes, therefore, that by explicating a concept in the way that Carnap suggests we merely block the conceptual imbalance, not resolve it.

Now, there are various ways the conceptual engineer might respond to Strawson's objection, but I flag only a few of them here, for two reasons.[56] For one, I simply do not have space to do each response justice. For another, my particular case of explication bypasses Strawson's objection in a way that other cases of explication perhaps do not, and so it is unnecessary for me to mount a full defence against it. In brief, here are three options for the conceptual engineer:

(i) It is not entirely clear why Strawson thinks that the explication of a concept cannot tell us why use of the concepts originally involved in some philosophical problem lead us astray in the first place. If an explication can tell us this, his argument does not necessarily go through, at least not for all explications.[57]
(ii) Strawson bases his objection on what is arguably a very narrow conception of philosophical problems. If we construe philosophical problems in a different or

[55] Adapted from Pinder's (2020b) recent construal of Strawson's critique of Carnapian explication.

[56] See Cappelen (2020b), Pinder (2020b), and Simion & Kelp (2020) for such responses.

[57] See e.g., Pinder (2000b).

broader way, his argument does not necessarily go through, at least not for all explications.[58]

(iii) We might think that turning our back on certain philosophical problems (whatever shape or form) is not such a bad thing – perhaps it is exactly what needs to be done in order to make progress. If so, his argument does not necessarily go through, at least not for all explications.[59]

It should be clear, then, that although some explicators may have some explaining to do, Strawson has not sounded the death knell for explicators across the board. With that, let us turn to my particular case of explication and, more specifically, the way in which it bypasses Strawson's objection. Frist, as Deutsch makes clear, my aim as a conceptual engineer is rather modest, for my aim is to explicate the term 'assertion' as a philosophical term of art, not to explicate it as a folk term. Thus, as mentioned above, I do not have to manage all the pre-theoretical concerns that arise with the explication of a concept like KNOWLEDGE. Second, I am not using my explication as such to solve a philosophical problem. Instead, it is the theoretical elaboration of my candidate explicatum which is designed to solve philosophical problems, and, strictly speaking, that is distinct from my stipulated candidate explicatum. In other words, my particular case of explication cannot be charged with blocking a conceptual imbalance instead of resolving one, since I do not profess to solve a conceptual imbalance *qua* explication. Third, it might be argued that I am turning my back on the philosophical problem of the nature of assertion, in that, I claim that what is needed is a proposal for how the term 'assertion' might fruitfully be applied, rather than asking, 'What is assertion?', as though it is a difficult question about the underlying nature of some determinate philosophical category. If so, following a similar line to Simion & Kelp (2020), I would respond that I am merely turning my back on a pseudo-philosophical problem, which is no bad thing. Indeed, I think it is precisely what is needed in order to make progress on the topic of assertion. Fourth, I arrived at my candidate explicatum in a

[58] See e.g., Pinder (2000b)

[59] See e.g., Simion & Kelp (2020)

principled way, namely, by using as my starting point the paradigm case of assertion and considering the philosophically interesting phenomena that relate to it. As we have seen, almost all the philosophically interesting phenomena that relate to the speech act can be established solely in terms of the paradigm case. Plus, there is no immediate reason to think that any philosophically interesting phenomena that is peripheral in the study of assertion cannot be accounted for in terms of the paradigm case. Thus, it seems my explication could not result in anything other than the preservation of topic.[60]

[60] Admittedly, whether certain apparent borderline cases of assertion are best understood as cases of assertion is yet to be seen. However, this simply means that there is room for the possibility that these borderline cases are best understood in terms of something other than assertion and so are themselves not on topic.

3. Putting the Epistemology of Testimony to Work

In this chapter, following the recommendation from the previous chapter that our inquiry should be guided by, connected to, and framed within broader philosophical theories, I call upon the epistemology of testimony literature to consider the ways in which the speech act of assertion can put an individual in a position to know the truth of what is asserted. As I explain, such considerations are apt to guide us in our inquiry and thus lay the groundwork for constructing a theoretical elaboration of our candidate explicatum, (AE). Although there are various ways in which assertion can put an individual in a position to know the truth of what is asserted, there is one particular way in which assertion can put an individual in a position to know the truth of what is asserted which I think illuminates the speech act as I understand it: 'knowledge transmission'. However, to fully understand this independently plausible notion from the epistemology of testimony literature, we must understand the nature of the interpersonal relation at play between speakers and hearers in such cases. And, to understand this, we must understand the conversational dynamics which underpin communicative exchanges generally. Unfortunately, there is some confusion in the epistemology of testimony literature about how to understand these things. The result of this is that I find myself in a predicament. I intend to construct a theoretical elaboration of our candidate explicatum in light of the notion of knowledge transmission. But I do not have an accurate account of the conversational dynamics which underpin communicative exchanges generally. As such, I do not have an accurate account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, means that I do not have a fully developed view of knowledge transmission in light of which I can construct a theoretical elaboration of our candidate explicatum. This sets us up for the ambitious task I undertake in the subsequent chapter.

3.1. **Untangling Assertion & Testimony**

'Testimony', as we saw in the previous chapter (§2.2), is a philosophical term of art, which is intended to capture the diverse array of cases in which an individual makes an utterance which may serve as a source of knowledge for another individual. Indeed, the term extends over different types of speech act, though it is unclear just how many different types of speech act it may extend over. Thus, the epistemology of testimony is concerned with the way in which testimony, in its various forms, facilitates the epistemic relationship between an individual and a fact (or true proposition). Specifically, it is concerned with the following two questions:

- How does testimony put an individual in a position to know the truth of what is testified?
- Under what conditions does testimony put an individual in a position to know the truth of what is testified?

My focus, however, is on the speech act of assertion as I understand it:

> (AE): Essentially, assertion is a communicative act which involves a speaker saying that *p* with the intention that the hearer acquire a belief that *p*.[61]

Assertion so understood, as we saw in §2.5, is co-extensive with the paradigm case of testimony. As such, I am only concerned with the way in which assertion, as the paradigm case of testimony, facilitates the epistemic relationship between an individual and a fact (or true proposition). Specifically, I am only concerned with the following two questions:

- How does assertion (as the paradigm case of testimony) put an individual in a position to know the truth of what is asserted?

---

[61] Once more, I flesh this out using Grice's (1957; 1969; 1975; 1989) framework of communication, in §4.2.

- Under what conditions does assertion (as the paradigm case of testimony) put an individual in a position to know the truth of what is asserted?

Why begin our inquiry with such questions? The answer to this can be given in general terms: because to understand any particular thing it is helpful to know what that thing does, and how and under what condition it does it. Such questions, then, are apt to guide us in our inquiry and thus lay the groundwork for constructing a theoretical elaboration of our candidate explicatum.

3.2. **The Variability of Assertion as a Source of Knowledge**

Traditionally, epistemologists answer questions on the source of testimonial knowledge from the point of view of one of two general positions, namely, 'reductionism' or 'anti-reductionism'.[62] According to reductionism, when a hearer acquires testimonial knowledge on the basis of a speaker's assertion: (R1) it is nondistinctive, in that, it is acquired solely by means of other fundamental sources of knowledge (e.g., perception, memory, reasoning); (R2) it is reason-dependant, in that, on top of comprehending the speaker as having made an assertion, the hearer must depend on additional positive reasons for believing what is asserted; and (R3) the hearer does not enjoy defeasible entitlement to believe what is asserted as true by default.[63] According to anti-reductionism, when a hearer acquires testimonial knowledge on the basis of a speaker's assertion: (AR1) it is distinctive, in that, it is not acquired solely by means of other fundamental sources of knowledge, but is itself a fundamental source of knowledge in its own right, which stands alongside other fundamental sources of knowledge; (AR2) it is reason-independent, in that, it is not the case

[62] The first reductionist account of testimony is usually attributed to Hume (1748), whereas the first anti-reductionist account of testimony is usually attributed to Reid (1764).

[63] (R2) may be disambiguated either in terms of 'global reductionism' or 'local reductionism'. Global reductionism is the view that the hearer must have additional positive reasons for believing what is asserted by speakers in general, whereas local reductionism is the view that the hearer must have additional positive reasons for believing what is asserted by a speaker in a particular case.

that on top of comprehending the speaker as having made an assertion the hearer must also have additional positive reasons for believing what is asserted; and (AR3) the hearer enjoys defeasible entitlement to believe what is asserted as true by default.[64] Although discussions about the (ir)reducibility of testimonial knowledge traditionally bundle these individual points together into these two general positions, they need not be bundled together in this way.[65] As such, some defend hybrid positions which fall between reductionism and anti-reductionism as they are traditionally construed.[66]

Now, I hold the view that in order to provide a complete epistemology of testimony we should accept that there are various ways in which assertion can put an individual in a position to know the truth of what is asserted.[67] This is different from adopting a hybrid view between reductionism and anti-reductionism, since the claim is not that there is only one way in which assertion can put an individual in a position to know the truth of what is asserted, one way which falls between reductionism and anti-reductionism. Rather, the claim is that there are various ways in which assertion can put an individual in a position to know the truth of what is asserted, some of which can be explained in reductionist terms and some of which can be explained in anti-reductionist terms, respectively.[68] For instance, I think it is possible for a hearer to come to know what a speaker asserts by treating the speaker's assertion in much the same way that we treat many other empirical findings as sources of knowledge. That is, I think that a hearer can observe the fact that a speaker made an assertion and, against their own background knowledge or beliefs, exercise their reason

---

[64] Note, there is an important difference between a hearer being psychologically disposed to believe what is asserted as true and a hearer enjoying defeasible entitlement to believe what is asserted as true by default. However, those who endorse anti-reductionism as it is traditionally construed generally endorse both positions.

[65] Greco (2012) provides a helpful exposition of the ways in which epistemologists of testimony answer questions on the source of testimonial knowledge.

[66] See, for example, Lackey (2008) and Faulkner (2011).

[67] See Wright (2018) for a similar view.

[68] I wish to stay neutral on the further claim that some cases might also be best explained in terms of some sort of hybrid position between reductionism and anti-reductionism as the positions are traditionally construed.

to make inferences as to whether what is asserted is true reliably. Here is such a case from Lipton (2007):

New Tyre

> I need a new tyre for my bike and I ask the person in the bike shop what type of tyre I should get. He says that the best bet for me is the Roadrage XT, because it is hard-wearing and has good puncture-resistance. Should I believe him? Well, why did he recommend that tyre? Perhaps only because it would make him the most profit. Or perhaps he genuinely believes that it is my best bet, but he is incompetent to judge, having been brainwashed by the Roadrage sales representative. On reflection, I decide that neither of these possibilities […] supports a very good explanation of what he said. After all, I have gone to this person for my bicycling needs for years, and he has always treated me well. He has plenty of other brands of tyre in stock, and some of them are considerably more expensive than the Roadrage. So I infer that […] what he said is true. (244)

For another, I think it is possible for a hearer to come to know what a speaker asserts by treating the speaker themselves in much the same way that we treat mechanical instruments as sources of knowledge, such as thermometers or fuel gauges. That is, I think that a hearer can assume a speaker is reliable in what they assert given that the speaker is situated such that they would not easily have asserted a falsehood on that occasion. Here is an extreme case from Lackey (2008):

Consistent Liar

> When Bertha was a teenager, she suffered a head injury while ice skating and, shortly afterwards, became quite prone to telling lies, especially about her perceptual experiences involving wild animals. After observing this behavior, her parents became increasingly distressed and, after consulting various psychologists and therapists, finally took her to see a neurosurgeon, Dr Jones. Upon examining her, Dr

> Jones noticed a lesion in Bertha's brain which appeared to be the cause of her behavior, and so it was decided that surgery would be the best option to pursue. Unfortunately, Dr Jones discovered during the surgery that he couldn't repair the lesion—instead, he decided to modify her current lesion and create another one so that her pattern of lying would be extremely consistent and would combine in a very precise way with a pattern of consistent perceptual unreliability. Not only did Dr Jones keep the procedure that he performed on Bertha completely to himself, he also did this with the best of intentions, wanting his patient to function as a healthy, happy, and well respected citizen. As a result of this procedure, Bertha is now—as a young adult—a radically unreliable, yet highly consistent, believer with respect to her perceptual experiences about wild animals. For instance, nearly every time she sees a deer, she believes that it is a horse; nearly every time she sees a giraffe, she believes that it is an elephant; nearly every time she sees an owl, she believes that it is a hawk, and so on. At the same time, however, Bertha is also a radically insincere, yet highly consistent, testifier of this information. For instance, nearly every time she sees a deer and believes that it is a horse, she insincerely reports to others that she saw a deer; nearly every time she sees a giraffe and believes that it is an elephant, she insincerely reports to others that she saw a giraffe, and so on.[29] [B]ecause of her consistency as both a believer and a liar, those around her do not have any reason for doubting Bertha's reliability as a source of information. (53--54).

Plus, I think that it is possible for a hearer to come to know what a speaker asserts in virtue of the modal profile of the case. That is, I think that a hearer can reliably form a belief based on a speaker's assertion because said belief could not easily have been false under the circumstances the hearer happens to be in. Here is such a case from Goldberg (2005):

Milk Carton

> Frank is a writer who spends his mornings in the kitchen, where he follows a mildly eccentric ritual. Every day after finishing his bowl of cereal at around dawn, he discards whatever milk is left, but places the empty carton back in the fridge, where

> it remains until noon, when he takes out the trash. Today, his sister Mary and nephew Sonny are visiting. Having had a tiring journey, they sleep in and join Frank in the kitchen in the late morning. Mary sees the milk carton in the fridge, assumes it is full, and on that basis tells Sonny that there is milk in the fridge. As it happens, there is indeed milk in the carton this morning, since Frank forgot to perform his ritual. Upon overhearing Mary's testimony, Frank is reminded of his blunder and realizes that he forgot to discard the milk this morning; at the same moment, Sonny, on the basis of his mother's testimony, forms the belief that there is milk in the fridge. Had Frank performed his ritual, he would have immediately corrected Mary and would have informed his visitors that the carton was empty. (Gelfert 2014: 156, which paraphrases and simplifies Goldberg 2005: 302)

Of course, in order to provide a complete epistemology of testimony, we would need to explain how and under what conditions assertion can put an individual in a position to know the truth of what is asserted in each different way. However, I am not in the business of providing a complete epistemology of testimony.[69] Instead, I am in the business of providing a theoretical elaboration of our candidate explicatum, (AE). As such, I am concerned with just one particular way – a way we have not yet addressed – in which assertion can put an individual in a position to know the truth of what is asserted: 'knowledge transmission'.[70] Why knowledge transmission? Because I think it is in light of this independently plausible notion from the epistemology of testimony literature in particular that a theoretical elaboration of our candidate explicatum is best constructed.

---

[69] See Wright (2018) who attempts to provide a complete epistemology of testimony along these lines.

[70] Although their individual views take varying forms, advocates of knowledge transmission include Welbourne (1986), Burge (1993), Dummett (1994), Moran (2005; 2018), Hinchman (2005), Audi (2006), Faulkner (2011), and McMyler (2011), to name a few.

3.3. **Knowledge Transmission**

Knowledge transmission is perhaps most helpfully introduced by analogy. Consider a chain of people who have been infected with a virus. Each person must have the virus in order to pass it to the next person, and moreover there must be at least one person who ultimately acquired the virus from another source. Analogously, each person in the chain of transmitting knowledge that *p* must know that *p* in order to pass it to the next person, and moreover there must be at least one person in the chain who ultimately acquired knowledge that *p* from another source (e.g., perception, memory, reasoning).[71] The central idea behind knowledge transmission, then, is that when a knowledgeable speaker makes an assertion that *p*, the hearer does not merely come to know what the speaker knows, but acquires the speaker's knowledge that *p*, in that, their knowledge that *p* is *grounded upon* the speaker's knowledge that *p*. Crucially, this idea is bound up with the further idea that a certain interpersonal relation must hold between the speaker and the hearer in order to successfully transmit knowledge. More specifically, the hearer must believe the speaker themselves, and thus rely on them to be knowledgeable in what they assert, in order to acquire the knowledge in question.[72] We might say, then, that knowledge transmission is the process by which a speaker's knowledge that *p* becomes a hearer's knowledge that *p* (i.e., the hearer's knowledge that *p* is grounded upon the speaker's knowledge that *p*), in virtue of the fact that the reason why the speaker asserted that *p* is that the speaker knows that *p*, and whereby the content so asserted, *p*, comes to be known by the hearer by way of the hearer believing the speaker themselves. To be clear, when a hearer acquires testimonial knowledge by means of

---

[71] This analogy is developed from Lackey (2008: 47), who describes knowledge transmission in terms of buckets of water being passed down a chain of people. I use the notion of a viral infection instead because it captures more accurately the notion of knowledge transmission, since when an individual passes on a viral infection, they do not lose the viral infection themselves, whereas when an individual passes on a bucketful of water, they lose the water they originally had in their possession. Accordingly, when a speaker transmits knowledge, they do not lose their knowledge.

[72] Anscombe (1979) provides what is arguably the first philosophical treatment of what it is to believe someone, especially a speaker.

a speaker's assertion in this particular way, it is distinctive, in that, it is not acquired solely by means of other fundamental sources of knowledge but is itself a fundamental source of knowledge in its own right, which stands alongside other fundamental sources of knowledge. In other words, it satisfies at least (AR1) of the anti-reductionist position as it is traditionally construed.[73]

Now, as I say, I think it is this independently plausible notion from the epistemology of testimony literature in particular that can illuminate the speech act of assertion as I understand it. However, to fully understand knowledge transmission we must understand the nature of the interpersonal relation at play between speakers and hearers in such cases. To understand this, though, we must understand the conversational dynamics which underpin communicative exchanges generally.

3.4. **The So-called 'Problem of Cooperation'**

According to Faulkner (2011), the conversational dynamics which underpin communicative exchanges generally institute a unique practical and theoretical difficulty for communicative exchanges the purpose or direction of which is to transmit knowledge specifically. He calls it the 'problem of cooperation'. This, for him, means that the nature of the interpersonal relation at play between speakers and hearers in cases of knowledge transmission must be understood in a certain way. Let me explain.

The apparent practical and theoretical difficulty the so-called 'problem of cooperation' is supposed to capture is this. When speakers and hearers enter into communicative exchanges

---

[73] As we will see below in §3.4, some think that it is possible to adopt a view of knowledge transmission which satisfies condition (AR1) of the anti-reductionist position as it is traditionally construed, but yet does not satisfy conditions (AR2) or (AR3). Instead, they adopt a view of knowledge transmission which satisfies condition (AR1) but which also satisfies conditions (R2) and (R3) of the reductionist position as it is traditionally construed. Thus, they adopt a view of knowledge transmission which falls between reductionism and ant-reductionism – a hybrid position. As we will see (in §3.5), however, such a view of knowledge transmission does not seem plausible.

generally, they normally do so with conflicting practical interests. Plus, the hearer may be aware of the speaker's practical interests and how they conflict with their own. In communicative exchanges the purpose or direction of which is to transmit knowledge specifically, speakers want to be believed whether they assert the truth or whether they are insincere (e.g., lying), whereas listeners want to believe speakers only if they assert the truth. This can be put in game-theoretic terms, as follows: the best outcome for the speaker is one in which they assert the truth in some situations and are insincere in others, but the hearer believes what they assert in all of them; the best outcome for the hearer is one in which the speaker always asserts the truth, and the hearer believes them only if they do so. Of course, it is possible that a speaker's practical interests and a hearer's practical interests might align in some cases. However, this is beside the point. Why? Because the point of the so-called 'problem of cooperation' is that there normally is a conflict. In other words, the idea is that asserting the truth is not the default position for speakers and thus it cannot be the case that the default position for hearers is to believe speakers. Given the conflicting practical interests between speakers and hearers in communicative exchanges generally, and thus in communicative exchanges the purpose or direction of which are to transmit knowledge specifically, for a hearer to believe what a speaker asserts without being aware of additional positive reasons for doing so is irrational and thus incompatible with acquiring knowledge. This is Faulkner's so-called 'problem of cooperation'. He maintains, however, that this practical and theoretical difficulty can be overcome if we concede that the nature of the interpersonal relation at play between speakers and hearers in cases of knowledge transmission is such that hearers, on top of comprehending speakers as having made an assertion, must be aware of additional positive reasons for believing what is asserted.

Since Faulkner is a proponent of knowledge transmission, he holds the view that testimonial knowledge acquired by means of a speaker's assertion in this particular way is distinctive, i.e., not acquired solely by means of other fundamental sources of knowledge but itself a fundamental source of knowledge in its own right, which stands alongside other fundamental sources of knowledge. However, as his proposed solution to the so-called 'problem of cooperation' shows, he also holds the view that testimonial knowledge acquired by means of a speaker's assertion in this particular way is reason-dependant, i.e., on top of

comprehending the speaker as having made an assertion, the hearer must have additional positive reasons for believing what is asserted. As such, he holds the view that the hearer does not enjoy defeasible entitlement to believe what is asserted as true by default. In other words, he holds a view of knowledge transmission which satisfies (AR1) of the anti-reductionist position as it is traditionally construed yet does not satisfy (AR2) or (AR3) of the anti-reductionist position as it is traditionally construed. Instead, he adopts a view of knowledge transmission which satisfies (AR1) but also (R2) and (R3) of the reductionist position as it is traditionally construed. Thus, he adopts a view of knowledge transmission which falls between reductionism and ant-reductionism – a hybrid position.

3.5. **Against the So-called 'Problem of Cooperation'**

Wright (2018), however, argues that those who, like Faulkner, operate on the assumption that they are at liberty to adopt a view of knowledge transmission which is inconsistent with anti-reductionism as it is traditionally construed are in fact mistaken.[74] According to him, proponents of knowledge transmission should subscribe to a view of knowledge transmission which is consistent with anti-reductionism as it is traditionally construed simply in virtue of being proponents of knowledge transmission, i.e., a view of knowledge transmission which satisfies (AR1), (AR2), and (AR3). His argument takes the form of the following dilemma:

Premise 1: There are proponents of knowledge transmission who adopt a view of knowledge transmission which is inconsistent with anti-reductionism as it is traditionally construed. They must either endorse or deny the claim that a listener can acquire knowledge via knowledge

[74] I should note that Wright does not use the term 'anti-reductionism' when he presents his position or argument. However, he nonetheless talks in terms of the content set out in conditions (AR1), (AR2), and (AR3). Thus, re-phrasing his position and argument in terms of anti-reductionism is unproblematic. In fact, it makes his position and argument much clearer for our purposes.

transmission only if the explanation for why the hearer believes what the speaker asserts is the fact that they have additional positive reasons for doing so.

Premise 2: If the claim is true, then the acquisition of knowledge via knowledge transmission is impossible.

Premise 3: If the claim is false, then the claim that a hearer's acquisition of knowledge via knowledge transmission depends on the fact that they have additional positive reasons for believing what the speaker asserts is itself a false claim.

Conclusion: Adopting a view of knowledge transmission which is inconsistent with anti-reductionism as it is traditionally construed is incompatible with the claim that assertion transmits knowledge.[75]

Taking premise 1 as given, Wright provides specific arguments in support of premise 2 and premise 3. Let us take them in turn. In support of premise 2, he argues that when a knowledgeable speaker makes an assertion, but the hearer believes what the speaker asserts on the basis of additional positive reasons for doing so, strictly speaking the hearer does not rely on the speaker's assertion. Why? Because, he claims, when the explanation for why the hearer believes what the speaker asserts is the fact that they are aware of positive reasons for doing so, the explanation for why the hearer believes what the speaker asserts cannot also be the fact that the speaker asserted it. In that case, the hearer does not acquire the speaker's knowledge (i.e., the hearer's knowledge is not grounded upon the speaker's knowledge) in the appropriate way. At best, the hearer merely comes to know what the speaker knows. Why is it, though, that a hearer who forms a belief on the basis of additional positive reasons is not related to the speaker's assertion in the appropriate way necessary for knowledge transmission? Because, according to Wright, for a fact to be an explanation for why an individual believes something involves that individual being responsive to that fact in their

[75] See Wright (2018: 35). Note, further to what was said in fn.74, I have changed some of the wording for our purposes, but the spirit of Wright's argument is the same.

believing it. As such, a hearer cannot be guided by both the fact that the speaker made an assertion and the fact that they have positive reasons for believing what the speaker asserts. If, he says, one of the facts had been otherwise – if either the hearer had been aware of reasons against believing what the speaker asserts, or if the speaker had lacked knowledge of what is asserted – then the hearer would have either believed or not believed what is asserted, being guided by one fact or the other. Wright notes that this is so even in a case where both facts are such that they each guide the hearer in the direction of believing what is asserted, respectively.[76] He maintains, therefore, that proponents of knowledge transmission who adopt a view of knowledge transmission which is inconsistent with anti-reductionism as it is traditionally construed must claim that a hearer acquiring knowledge via knowledge transmission depends on them being aware of positive reasons for doing so, but where the fact that they are aware of positive reasons must not be the explanation for why they believe what the speaker asserts. This, he claims, is an untenable position, which brings us to his argument in support of premise 3.

Recall that Faulkner's concern is that believing what a speaker asserts without being aware of additional positive reasons for doing so is irrational and thus incompatible with acquiring knowledge. Wright, however, makes the intuitive observation that there is no difference in terms of rationality between someone who believes something whilst being aware of positive reasons for doing so, but where these positive reasons do not explain why they

---

[76] Note, one of Wright's main sources of inspiration for this argument is Hinchman (2014), who holds that there is a sense in which knowledge transmission involves the speaker issuing the hearer with an invitation to trust, not only for the truth of what is asserted but to be knowledgeable on the topic in question, so that the hearer can close further inquiries into the matter. Accordingly, if the hearer refuses the speaker's invitation to trust, the hearer fails to acquire the speaker's knowledge. Wright's idea, then, is that when the explanation for why the hearer believes what is asserted is the fact that the speaker is aware of positive reasons for doing so, this is tantamount to refusing the speaker's invitation to trust, for the hearer does not believe what the speaker asserts in the way that the speaker intends. (I say more about the way that I understand the notions of trust and the invitation to trust in §4.3 and §4.3.2, respectively.)

believe that thing, on the one hand, and someone who believes something but is not aware of positive reasons for doing so, on the other. On the assumption that this intuitive observation is correct, he argues that if a hearer is aware of positive reasons for believing what a speaker asserts – but where these positive reasons are not the explanation for why the hearer believes what is asserted – then so too is a hearer who is not aware of positive reasons for believing what is asserted. He maintains, therefore, that it cannot plausibly be claimed that a hearer being aware of additional positive reasons is a necessary condition for them to acquire knowledge via knowledge transmission.

Intuitions notwithstanding, Wright concludes that proponents of knowledge transmission who operate on the assumption that they are at liberty to adopt a view of knowledge transmission which is inconsistent with anti-reductionism as it is traditionally construed are in fact mistaken. Instead, proponents of knowledge transmission should subscribe to a view of knowledge transmission which is consistent with anti-reductionism as it is traditionally construed simply in virtue of being proponents of knowledge transmission.

3.6. **Where Do We Go from Here?**

Given Wright's dilemma, I submit that Faulkner's so-called 'problem of cooperation' misconstrues the conversational dynamics which underpin communicative exchanges generally, and thus misconstrues the nature of the interpersonal relation at play in cases of knowledge transmission specifically. As I see it, the conversational dynamics which underpin communicative exchanges generally must be such that the nature of the interpersonal relation at play in cases of knowledge transmission specifically is such that the default position for speakers *is* to assert the truth. In that case, hearers *do* enjoy a defeasible entitlement to believe speakers, and thus what is asserted as true, by default. Accordingly, it *is not* the case that on top of comprehending the speaker as having made an assertion the hearer need be aware of additional positive reasons for believing what is asserted. In other words, I think the so-called 'problem of cooperation' should be dissolved in such a way that supports a view of knowledge transmission which is consistent with anti-reductionism as it

is traditionally construed (i.e., a view of knowledge transmission which satisfies (AR1), (AR2), and (AR3)).

This, however, puts me in a predicament. I intend to construct a theoretical elaboration of our candidate explicatum in light of the notion of knowledge transmission. But I do not yet have an accurate account of the conversational dynamics which underpin communicative exchanges generally. Thus, I do not have an accurate account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This means that I do not have a fully developed view of knowledge transmission in light of which I can construct a theoretical elaboration of our candidate explicatum. What I need to do, then, is construct my own account of the conversational dynamics which underpin communicative exchanges generally. I can then use this to construct my own account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission in light of which I can construct my theoretical elaboration of our candidate explicatum. Indeed, this is the ambitious task I take up in the next chapter.

3.7. **Appendix B: safeguarding knowledge transmission**

It would be remiss of me not to mention that the notion of knowledge transmission has come under attack in the epistemology of testimony literature, most famously from Lackey (2008). Arguably, her most serious attack takes the form of her cases of so-called 'selfless assertions', the definition for which is as follows:

> A speaker's assertion is 'selfless' iff:
>
> (i) for purely non-epistemic reasons, the speaker does not believe that *p*
> (ii) despite this lack of belief, the speaker is aware that *p* is very well supported by all of the available evidence
> (iii) because the speaker is aware that *p* is very well supported by all of the available evidence, they assert that *p*.

Here is her most well-known case of a so-called 'selfless assertion':

Creationist Teacher

> Stella is a devoutly Christian fourth-grade teacher, and her religious beliefs are grounded in a deep faith that she has had since she was a very young child. Part of this faith includes a belief in the truth of creationism and, accordingly, a belief in the falsity of evolutionary theory. Despite this, she fully recognizes that there is an overwhelming amount of scientific evidence against both of these beliefs. Indeed, she readily admits that she is not basing her own commitment to creationism on evidence at all but, rather, on the personal faith that she has in an all-powerful Creator. Because of this, Stella does not think that religion is something that she should impose on those around her, and this is especially true with respect to her fourth-grade students. Instead, she regards her duty as a teacher to involve presenting material that is best supported by the available evidence, which clearly includes the truth of evolutionary theory. As a result, after consulting reliable sources in the library and developing reliable lecture notes, Stella asserts to her students, "Modern-day *Homo sapiens* evolved from *Homo erectus*," while presenting her biology lesson today. Though Stella herself neither believes nor knows this proposition, she never shares her own personal faith-based views with her students, and so they form the corresponding true belief solely on the basis of her reliable testimony. (48).

Lackey's argument takes the following form:

Premise 1: In cases of so-called 'selfless assertion' (in this instance, *Creationist Teacher* specifically), the hearers come to know what the speaker asserts by believing the speaker's assertion.

Premise 2: If the knowledge that the hearers acquire by believing the speaker's assertion cannot be accounted for in terms of knowledge transmission, then knowledge transmission is false.

Premise 3: The knowledge that the hearers acquire by believing the speaker's assertion cannot be accounted for in terms of knowledge transmission.

Conclusion: Knowledge transmission is false.

Now, there have been various responses to *Creationist Teacher* and other such cases of so-called 'selfless assertions' in the epistemology of testimony literature in defence of knowledge transmission, and I am afraid I do not have anything particularly novel to add. However, I will make a few comments.

First, since I think there are various ways (other than knowledge transmission) in which the speech act of assertion can put an individual in a position to know the truth of what is asserted, it is an open possibility (for me, at least) that *Creationist Teacher* can be explained in terms of one of these instead. For instance, perhaps the case is best understood as one in which the students treat Stella in much the same way that they treat mechanical instruments as sources of knowledge, merely as a reliable gauge of the truth. On this interpretation, premise 2 of Lackey's argument is false. (Thank you to Stephen Wright (personal correspondence) for a helpful conversation about this.)

Second, although Lackey claims that Stella makes an assertion to her students, this is perhaps not an entirely plausible claim. Pedagogically speaking, it certainly sounds odd (if not worrying) to claim that schoolteachers are in the business of asserting truths to their students, rather than, say, presenting them with content from a syllabus which has been deemed appropriate/correct by the institution (i.e., the school and its governing bodies) rather than strictly speaking true. Indeed, institutions and their schoolteachers are usually rather sensitive to the way in which they present content to students given that students come from various cultural, religious, and ethnic backgrounds. The role of a schoolteacher is not to indoctrinate, after all. Moreover, since students are often taught about conflicting views and positions during the course a lecture/seminar (let alone the course of a module), it is not at all obvious how students would be expected to respond epistemically or otherwise to a teacher who is making so many conflicting assertions to them. The speech act Stella performs might thus be best understood not as an assertion but as something else, like

'presenting'. On this interpretation, premise 1 of Lackey's argument is false. (See Bourne, Caddick Bourne, & Jarmy (2016) and Milić (2017) for the view that teaching does not involve asserting truths.)

Third, even if we do accept Lackey's claim that an assertion that modern day Homo sapiens evolved from Homo erectus has been made, this assertion is perhaps best described as being made by the institution (i.e., the school and its governing bodies), not Stella. Stella may thus be seen as an appendage of the institution – a 'mouthpiece', if you will. As such, it does not matter whether she knows that modern day Homo sapiens evolved from Homo erectus or not, because the institution knows and asserts this, the students believe the institution, and thus knowledge is transmitted (albeit by proxy through Stella). On this interpretation, premise 3 of Lackey's argument is false. (See Faulkner (2011) and Burge (2013) for a similar sort of view.) (Note, I see no reason to worry about the suggestion that institutions can have knowledge and make assertions, even when certain individuals who make up those institutions have conflicting beliefs. Indeed, we regularly talk about institutions having beliefs, knowledge, values, etc., as well as performing speech acts. Consider, for instance, BP's apology for the oil spill which polluted the Gulf of Mexico. For more on assertions made by institutions (or 'group assertions'), see Hughes (1984), Meijers (2007), Tollefsen (2007; 2009; 2011; 2020), (Miranda) Fricker (2012), Lackey (2015; 2018), and Faulkner (2018).)

Fourth – and certainly the most bullet-biting response I can think of in response to so-called 'selfless assertions' generally – since such assertors are so bizarre in their epistemic profile, we might simply hold that hearers at best merely come to acquire a (true) belief on the basis of such assertions, not knowledge. Why? Because even though what the speaker asserts is supposedly very well supported by all the available evidence, we have the complicating factor that they themselves do not believe what they assert, which (even by Lackey's standards) acts as a defeater against all the available evidence that the speaker supposedly has. On this interpretation, premise 1 of Lackey's argument is false. (See Audi (2006) for this approach.)

In sum, whether or not we want to say that *Creationist Teacher* (or indeed any case of so-called 'selfless assertion') can be explained in terms of knowledge transmission, it is not obvious that it amounts to a counterexample to knowledge transmission.

4. A Gricean Account of Assertion

In this chapter, I construct my theoretical elaboration of our candidate explicatum, (AE). In order to do this, I construct my own account of the conversational dynamics which underpin communicative exchanges generally. I then use this to construct my own account of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which satisfies (AR1), (AR2), and (AR3) of the anti-reductionist position as it is traditionally construed – in light of which I can construct my theoretical elaboration of our candidate explicatum. As I see it, Grice's (1957; 1969; 1975; 1989) framework of communication is best placed to help undertake this ambitious task, so I draw on it liberally, as well as revise and develop elements of it. I should add that my approach in this chapter is programmatic, and so I do not shy away from showing my working or flagging areas of uncertainty. Nonetheless, by the end of the chapter, I hope to have shown that (AE), and indeed my theoretical elaboration of it, can fruitfully be applied to almost all the philosophically interesting phenomena related to the speech act (i.e., items (1) through to (10) on the substantive list – I save the final item on the substantive list, item (11), for the subsequent chapter), as well as leave open the possibility that it may also fruitfully be applied to any outstanding philosophically interesting phenomena peripheral in the study of the speech act. I start by setting out my general epistemological commitments.

4.1. **Epistemological Framework**

Williams (1973) famously claims that belief aims at truth. He bases this claim on the following observations:

- Truth and falsehood are appropriate dimensions of assessment of beliefs (as opposed to other psychological states and dispositions)
- To believe that *p* just is to believe that *p* is true. (137)

How, though, should the thought that belief aims at truth be understood? An increasingly popular view is that beliefs are governed by a so-called 'epistemic norm', namely, a so-called 'truth norm', according to which the very nature of belief is that it is, in some sense, constituted by its being so governed, or that its being so governed is, in some sense, essential to it.[77] The central idea behind this view is that the so-called 'truth norm' has prescriptive force. That is, it indicates what ought to be done. However, this is not the only way of understanding the thought that 'belief aims at truth'. Although I agree that belief aims at the truth in a way that other attitudes do not, I understand this in terms of biological design rather than so-called 'epistemic norms'. From the point of view of biological design, belief has a different and specialised function from, say, desire. Papineau (2013) is instructive here:

> Desires relate to results. Each desire type has the function of generating actions that will lead to some specific outcome, such as food, or fine wine, or watching Tottenham Hotspur play football. By contrast, beliefs have no results to call their own. Their function is not to produce specific results, but to help whichever desires are active to select those actions that will conduce to their satisfaction. To do this, beliefs need to carry [true] information about the environment, [true] information that is relevant to which actions will produce which results. (72–74)

Of course, some might wish to talk in terms of biological 'norms', but it is important that we do not misconstrue such biological 'norms' as having genuinely prescriptive force. How, though, do we acquire true beliefs while also avoiding error? My answer to this is reliabilist in nature, namely, by bringing it about that the processes by which we acquire beliefs are truth conducive.[78] Indeed, I think that knowledge in general is the product of forming true beliefs in this way. That said, it seems somewhat artificial to talk purely in terms of reliable

---

[77] For views in this vicinity see Railton (1994), Wedgwood (2000; 2007), Velleman (2000), Brandom (2001), and Whiting (2010).

[78] See Goldman (1979; 1986)

processes without mentioning the role that evidence plays in such processes.[79] As such, I hold the general view that an individual's beliefs must accord with their evidence in a reliable way to amount to knowledge, i.e., where an individual's beliefs accord with their evidence if and only if that evidence is reliably connected to the truth of those beliefs.[80] Furthermore, I incorporate into my reliabilism a brand of contextualism based on Lewis (1996), who writes:

> We have all sorts of everyday knowledge, and we have it in abundance […] Besides knowing a lot that is everyday and trite, I myself think that we know a lot that is interesting and esoteric and controversial. We know a lot about things unseen: tiny particles and pervasive fields, not to mention one another's underwear […] But on these questions, let us agree to disagree peacefully with the champions of 'post-knowledgeism'. The most trite and ordinary parts of our knowledge will be problem enough. For no sooner do we engage in epistemology – the systematic philosophical examination of knowledge – than we meet a compelling argument that we know next to nothing. (549)

The 'compelling argument' Lewis is referring to is the well-known sceptical argument, which may take the following form:

> Premise 1: If I know that I have hands, then I know that I am not a brain in a vat.
> Premise 2: I do not know that I am not a brain in a vat.
> Conclusion: I do not know that I have hands.

---

[79] I use 'evidence' very broadly here, where anything from everyday perceptual experiences to comprehending speakers as making assertions may be understood as a form of evidence. (I am aware that some (e.g., Moran 2005; 2018) think that assertions should not strictly speaking be understood in terms of evidence. But I take it that their use of the term 'evidence' is considerably narrower than mine. As such, in real terms, I am not in opposition to such positions.) Anyway, my meaning should become clear as we work through the chapter.

[80] See Comesaña (2010), Goldman (2011), and Goldberg (2018).

Essentially, the problem is this. It seems that knowledge must be by definition infallible. However, if it is claimed that an individual knows that *p*, and yet it is also granted that said individual cannot eliminate a certain possibility in which not-*p*, it seems that individual does not know that *p*, after all. In other words, speaking of fallible knowledge, of knowledge despite uneliminated possibilities of error, sounds contradictory. Lewis pitches this as a dilemma:

> [W]e know a lot; knowledge must be infallible; yet we have fallible knowledge or none (or next to none). We are caught between the rock of fallibilism and the whirlpool of skepticism. Both are mad! (550)

He attempts to resolve this dilemma by employing both relevant alternatives theory and contextualism, granting the infallibilist the condition that knowledge requires the elimination of every relevant alternative but construing this condition in a context-sensitive way, whereby sceptical possibilities are ordinarily irrelevant. The idea behind this is that, in 'ordinary contexts', we do not consider sceptical possibilities and so these possibilities are 'properly ignored', and, as such, we have knowledge. But, in 'extraordinary contexts', when sceptical possibilities come to our attention and so are not 'properly ignored', we do not have knowledge. He writes:

> Maybe epistemology is the culprit. Maybe this extraordinary pastime robs us of our knowledge. Maybe we do know a lot in daily life; but maybe when we look hard at our knowledge, it goes away. But only when we look at it harder than the sane ever do in daily life; only when we let our paranoid fantasies rip. That is when we are forced to admit that there always are uneliminated possibilities of error, so that we have fallible knowledge or none. (Ibid.)

Strictly speaking, then, Lewis offers an account of knowledge *ascriptions*. That is, he does not offer an account of knowledge as such, i.e., when *S* knows that *p*, but rather an account of sentences that ascribe knowledge, i.e., when ascriptions of the form '*S* knows that *p*' are true. His concern is thus metalinguistic. Specifically, for Lewis, 'know' is a type of context-

sensitive term, such that a sentence containing it will express distinct propositions in different conversational contexts. So, when he says that epistemology robs us of our knowledge, what he means is that engagement in epistemology can create a type of extraordinary context in which '*S* knows that *p*' is false when it would otherwise be true in an ordinary context. With this, Lewis offers his relevant alternatives theory for knowledge ascriptions:

> A subject *S* satisfies '*S* knows that *p*' in context *C* if and only if *S*'s evidence eliminates every not-*p* possibility, except for those that are properly ignored in *C*.[81]

He then combines his relevant alternatives theory with a detailed account of contextualism, which consists of a set of rules designed to show when a possibility in a context is or is not 'properly ignored', and thus determine the relevant possibilities in that context. His rules are as follows:

> Rule of actuality: a possibility that actually obtains for the subject is always relevant and thus not properly ignored.
>
> Rule of belief: a possibility that the subject believes or should believe obtains is always relevant and thus not properly ignored.
>
> Rule of resemblance: any possibility that saliently resembles a relevant possibility (made relevant by any rule other than the rule of resemblance itself) is always relevant and thus not properly ignored.
>
> Rule of attention: any possibility that is being attended to is always relevant and thus not properly ignored.

---

[81] Adapted for simplicity from Lewis (1996: 554).

> Rule of reliability: possibilities concerning errors in reliable processes are defeasibly irrelevant and thus (properly) ignorable.
>
> Rule of method**:** possibilities concerning errors in inference are defeasibly irrelevant and thus (properly) ignorable.
>
> Rule of conservatism: possibilities that are normally ignored are defeasibly irrelevant and thus (properly) ignorable. (554–560)[82]

To clarify, the idea behind the distinction between eliminating a possibility and (properly) ignoring a possibility is this. Although an individual's evidence may not eliminate a certain possibility, the context may still make such a possibility ignorable. That is, although a certain possibility will be left uneliminated by an individual's evidence, it does not need to be eliminated by their evidence, because the context is such that it never really figures as a possibility in the first place. Indeed, some possibilities never can be eliminated by an individual's evidence, only ignored. Notice that the notion of eliminating, unlike the notion of ignoring, is conceptually tied to the notion of evidence. That is, it is in virtue of an individual's evidence that possibilities are eliminated, whereas no evidence is needed to ignore possibilities. Cases of properly ignoring, then, are those cases in which a possibility is ignored in line with Lewis's rules. As an illustration, consider the following from Lewis:

> What (non-circular) argument supports our reliance on perception […]? My visual experience, for instance, depends causally on the scene before my eyes, and what I believe about the scene before my eyes depends in turn on my visual experience. Each dependence covers a wide and varied range of alternatives. Of course, it is possible to hallucinate – even to hallucinate in such a way that all my perceptual

[82] Admittedly, there may be questions and apparent problems which arise for Lewis's rules, but I do not address them here. For recent endorsements and developments of Lewis's contextualism, see Blome-Tilman (2014) and Ichikawa (2017).

> experience and memory would be just as they actually are. That possibility never can be eliminated. But it can be ignored. And if it is properly ignored – as it mostly is – then vision gives me knowledge. (558)

Lewis is referring here to the well-known idea that there can be no non-circular confirmation of reliability. However, he does not view this as a vicious circularity, but rather as a legitimate form of epistemic bootstrapping. Thus, we can see his thinking behind the rule of reliability and the rule of method:

> Within limits, we are entitled to take [reliable processes] for granted. We may properly presuppose that they work without a glitch in the case under consideration. Defeasibly – very defeasibly! – a possibility in which they fail may properly be ignored […] We do not, of course, presuppose that nowhere ever is there a failure of, say, vision. The general presupposition that vision is reliable consists, rather, of a standing disposition to presuppose, concerning whatever particular case may be under consideration, that we have no failure in that case. (Ibid.)

In sum, Lewis's account of knowledge ascriptions is the result of conjoining his relevant alternatives theory with his contextualism, as follows:

> A subject *S* satisfies '*S* knows that *p*' in context *C* if and only if *S*'s evidence eliminates every not-*p* possibility, except for those that are properly ignored in *C*, where the relevance of a possibility in *C* is determined by all and only the rules of actuality, belief, resemblance, reliability, method, conservatism, and attention.

I think that Lewis's picture is fundamentally correct. However, two brief comments are in order. First, Lewis does not think that belief is required to satisfy '*S* knows that *p*'. His reason for this is based on a case which purports to show that a timid student who knows the answer to a teacher's question but has no confidence that they are right does not believe what they know. Armstrong (1973), however, suggests a different interpretation of the case, namely, that the student knows and believes the answer to the teacher's question, but they

do not know or believe that they know and believe the answer to the teacher's question. Armstrong's interpretation not only makes better sense of the timid student case but also safeguards the intuitive thought that belief is required to satisfy '*S* knows that *p*'. Accordingly, I amend Lewis's picture to include a belief condition.[83] Second, Lewis does not think that justification is required to satisfy '*S* knows that *p*'. The notion of justification he rejects, however, is that of an internalist nature, according to which an individual must have access to what supports their belief in order to have knowledge, which may involve the individual being aware (or capable of being aware) of positive reasons that make their belief rational or being able to provide positive reasons for their belief. Given my reliabilist sensibilities, I agree with Lewis that justification of this nature is not necessary for knowledge. Nonetheless, we may still understand Lewis as thinking that a reliabilist conception of justification, or 'warrant', is the mark of knowledge, since this does not necessarily involve the individual being aware (or capable of being aware) of certain facts that make their belief that *p* rational or being able to provide reasons for their belief that *p*.[84] With this laid out, I propose the following:

'*S* knows that *p*' is true in context *C* iff:

(i) *S* believes that *p*
(ii) *S*'s belief that *p* is formed in a reliable way, and
(iii) *S*'s evidence eliminates every not-*p* possibility, except for those not-*p* possibilities that are properly ignored in *C*, where the relevance of a possibility in *C* is determined by all and only the rules of actuality, belief, resemblance, reliability, method, conservatism, and attention.[85]

---

[83] Note, Blome-Tilman (2014) and Ichikawa (2017) both add a belief condition to their respective versions of Lewis's contextualism.

[84] Following Mellor (1988), I use 'warrant' instead of 'justification'. 'Warrant', unlike 'justification', does not have strong internalist connotations, which fits better with my externalist sensibilities.

[85] Although I think that Lewis's contextualism is fundamentally correct, my account of assertion does not depend on Lewis's contextualism. Moreover, my account of assertion, as presented here, should not be taken as an

Now, there is of course a lot more to say in filling out and defending my general epistemological framework. However, that is not my aim here. I am merely marking out in broad strokes my general epistemological commitments. What is important to note, though, is that since I think knowledge in general is the product of forming beliefs through reliable processes, I take assertion to be just another instance of that. And I think there are epistemologically significant differences in the ways in which that reliability can be realised by means of assertion alone, as I mentioned n the previous chapter (§3.2):

- It is possible for a hearer to come to know what a speaker asserts by treating the speaker's assertion in much the same way that we treat many other empirical findings as sources of knowledge. That is, a hearer can observe the fact that a speaker made an assertion and, against their own background knowledge or beliefs, exercise their reason to make inferences as to whether what is asserted is true reliably.

- It is possible for a hearer to come to know what a speaker asserts by treating the speaker themselves in much the same way that we treat mechanical instruments as sources of knowledge, such as thermometers or fuel gauges. That is, a hearer can assume a speaker is reliable in what they assert given that the speaker is situated such that they would not easily have asserted a falsehood on that occasion.

- It is possible for a hearer to come to know what a speaker asserts in virtue of the modal profile of the case. That is, a hearer can reliably form a belief based on a

---

argument for Lewis's contextualism as such, though if it can be shown to function as one, that is fine with me. The key point is this: thinking in terms of Lewis's notions of eliminating possibilities and (properly) ignoring possibilities – and understanding these notions in terms of the rules that he provides – is helpful for articulating my account of assertion. It is of course possible, however, to adopt Lewis's notions of eliminating possibilities and (properly) ignoring possibilities – and understanding these notions in terms of the rules (or at least most of the rules) that he provides – without endorsing his contextualism. In other words, they are in principle available to the subject-sensitive invariantist and the classical invariantist as well. So, if someone wishes to endorse my view of assertion without endorsing Lewis's contextualism, they are welcome to.

speaker's assertion because said belief could not easily have been false under the circumstances the hearer happens to be in.

However, as I said, there is one particular way in which that reliability can be realised by means of assertion alone which I think illuminates the speech act, as follows:[86]

- It is possible for a hearer to come to know what a speaker asserts by means of knowledge transmission. That is, a speaker's knowledge that *p* may become a hearer's knowledge that *p* (i.e., the hearer's knowledge that *p* may be grounded upon the speaker's knowledge that *p*), in virtue of the fact that the reason why the speaker asserted that *p* is that the speaker knows that *p*, and whereby the content so asserted, *p*, comes to be known by the hearer by way of the hearer believing the speaker themselves.

With my general epistemological framework set out, I am now ready to introduce Grice's framework of communication, with which I will construct my account of the conversational dynamics which underpin communicative exchanges generally. I will then use this to construct my account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as the position is traditionally construed – in light of which I can construct my theoretical elaboration of our candidate explicatum.

[86] Although I do not pursue it here, by constructing a theoretical elaboration of our candidate explicatum in light of knowledge transmission, it is likely that we will come to better understand cases in which reliability is realised by means of assertion alone, though not by means of knowledge transmission specifically.

4.2. **Grice's Framework of Communication**

Grice's (1957; 1969; 1975; 1989) work in the philosophy of language may be seen as comprising two unrelated theories: the 'theory of meaning' and the 'theory of conversation'. The former traditionally concerns whether sentence-meaning and word-meaning within a language can be analysed in terms of what a speaker means on an occasion of use, where what a speaker means on an occasion of use is analysed in terms of the speaker's intentions (what Grice calls 'non-natural meaning'). The latter traditionally concerns the way in which a speaker may mean something over and above what they have uttered in virtue of exploiting certain communicative standards (what Grice calls 'implicature'). I, however, do not align myself with the view that these two theories are unrelated. My sympathies lie with Neale (1992) on this point:

> It is at least arguable that the Theory of Conversation is a component of the Theory of Meaning. And even if this interpretation is resisted, it is undeniable that the theories are mutually informative and supportive, and that they are of more philosophical, linguistic, and historical interest if the temptation is resisted to discuss them in isolation from one another. (512)

As I see it, although Grice's theory of meaning is traditionally concerned with developing a so-called 'intention-based semantics', and although his theory of conversation is traditionally concerned with elucidating the so-called 'semantic/pragmatic distinction', the interest and importance of these theories is not exhausted by such concerns. What Grice provides us with, when we unite his theory of meaning with his theory of conversation, is a framework of communication.

The mark of successful communication is a reciprocal relationship between the speaker and the hearer, where each are aware of certain psychological states in the other. Grice's framework of communication captures precisely this picture by focusing on a specific type

of case, one in which a speaker informs a hearer of the facts, namely, 'telling'.[87] According to him, the relevant psychological states, in cases of telling, are beliefs and intentions. Specifically, the speaker exercises intentions to influence the hearer's beliefs in a certain way. He illustrates this with what I will call the 'telling mechanism':

> Telling mechanism: *A* intends that *B* acquire a belief that *p* on the basis of *B* recognising *A*'s intention that *B* acquire a belief that *p*.[88]

The idea is that in order to comprehend a speaker's act of telling the hearer must assume that the speaker satisfies this mechanism. Furthermore, assuming that the speaker satisfies this mechanism involves employing the expectation that the speaker adheres to certain standards which hold during our communicative exchanges generally.[89] Grice calls this the 'co-operative principle':

> Co-operative principle: Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (1989: 26).

---

[87] Grice (1957: 44) writes: 'What we want to find is the difference between, for example, 'deliberately and openly letting someone know' and 'telling' and between 'getting someone to think' and 'telling''.

[88] See Grice (1957). Note, Grice later changes his mind about the aim of the intention. He writes:

> I wish to regard the M-intended effect common to indicative-type utterances as being, not that the hearer should believe something (though there is frequently an ulterior intention to that effect) but that the hearer should *think that the utterer believes* something. (1989: 123)

However, as mentioned in the previous chapter (§2.5), I think this is the wrong way to go.

[89] For a helpful exposition of Grice's framework of communication as well an interesting philosophical treatment of the ways in which speakers deliberately engage in unsuccessful communication, see Bourne and Caddick Bourne (2018). (Note, I discuss Bourne and Caddick Bourne's work in relation to so-called 'bald-faced lies' in §5.2.6.)

He proposes that four categories of maxims fall within this general principle – 'quality', 'quantity', 'relation', and 'manner' – under which fall certain maxims:

Maxims of quality

The super-maxim:

- Try to make your contribution one that is true.

Specific maxims:

- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

Maxims of quantity

- Make your contribution as informative as is required (for the current purposes of the exchange).
- Do not make your contribution more informative than is required.

Maxims of relation

- Be relevant.

Maxims of manner

The super-maxim:

- Be perspicuous.

Various maxims:

- Avoid obscurity of expression.
- Avoid ambiguity.
- Be brief.
- Be orderly. (26–27)

Although the co-operative principle and maxim categories hold during our communicative exchanges generally, the maxims themselves do not, since Grice formulates these with a specific type of communicative exchange in mind:

> The conversational maxims […] are specially connected (I hope) with the particular purposes that talk (and so, talk exchange) is adapted to serve and is primarily employed to serve. I have stated my maxims as if this purpose were a *maximally effective exchange of information*; this specification is, of course, too narrow, and the scheme needs to be generalized to allow for such general purposes as influencing or directing the actions of others. (Grice 1989: 28 – emphasis added)

A maximally effective exchange of information is a communicative exchange the purpose or direction of which is to inform each other of the facts, by means of the telling mechanism. This reveals itself when we look more closely at the maxims of quality in particular:

Maxims of quality

The super-maxim:

- Try to make your contribution one that is true.

Specific maxims:

- Do not say what you believe to be false.

- Do not say that for which you lack adequate evidence.

Notice that the supermaxim features the term 'contribution', and the specific maxims feature the term 'say'. (By unpacking Grice's notion of 'saying' first, his notion of a 'contribution' will become clearer.) For Grice, 'saying' is a technical term:

> In the sense in which I am using the word *say*, I intend what someone has said to be closely related to the conventional meaning of the words (the sentence) he has uttered. (25)

Importantly, however, he adds:

> I want to say that (1) "*U* (utterer) said that *p*" entails (2) "*U* did something *x* by which *U* [non-naturally] meant that *p*". (87)

For Grice, then, when a speaker says that *p*, not only do they make an utterance which expresses *p*, but, in virtue of making said utterance, they intend that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*. In other words, his notion of saying incorporates his telling mechanism. His notion of saying, therefore, is more complex than Austin's, in that, it does not merely denote a locution but denotes an illocution built on top of a locution.[90] Neale (1992) is instructive here:

> Anyone who reads 'Logic and Conversation' and 'Further Notes' in isolation from the rest of the William James Lectures, is almost certain to miss the relationship that

[90] Much has been said about Grice's notion of 'saying', both with respect to his account of proposition expression and his claim that saying necessarily involves telling. The former is generally criticised for being insufficient, with the issue being that 'what is said' cannot alone be established by reference to the conventional meaning of the words (the sentence) uttered. The latter is generally criticised for being unnecessarily strong, with the issue being that speakers can say that *p* without telling that *p*.

> Grice sees between utterers' intentions and what is said. For something to be (part of) what *U* says on a particular occasion, it must also be (part of) what *U* meant, i.e., it must be backed by a complex intention of the sort that forms the backbone of Grice's Theory of Meaning. If *U* utters the sentence "Bill is an honest man" ironically, on Grice's account *U* will *not* have said that Bill is an honest man: *U* will have *made as if to say* that Bill is an honest man. (523)

Following this, Grice claims that, if a speaker makes an utterance which is seemingly uncooperative and does so openly (i.e., by 'exploiting' the conversational maxims), then the speaker may be understood as telling the hearer something (i.e., satisfying the telling mechanism) without actually 'saying' it, but rather by 'implicating' it. The idea is that what a speaker implicates is whatever is needed to reconcile the speaker's apparent uncooperativeness with the expectation that they are in fact cooperative. This can be illustrated using Grice's well-known referee example. Suppose a referee writes a reference for a candidate for an academic position but only utters, 'He has very neat handwriting'. Given that the reference is supposed to focus on academic expertise, the information that the referee provides seems to be irrelevant, as well uninformative. In order to reconcile the speaker's utterance with the maxims of relation and the maxims of quantity, the hearer expects that the referee is being as relevant and as informative as it is possible to be within the remit of the task, which, in this case, is to provide positive information about the candidate. So, the hearer calculates that the referee has used their utterance of 'He has very neat handwriting' to implicate that the candidate is not suitable for the position.[91] Notice,

---

[91] Grice introduces three different types of implicature as part of his framework of communication: 'conversational implicature', 'conventional implicature', and 'scalar implicature'. In this thesis, I only discuss conversational implicature. (Thus, from now on, whenever I use the term 'implicature', I should be understood as referring to conversational implicatures only.) Note, one of the key features of conversational implicatures is that they can be 'cancelled'. This is best explained by example. Suppose again that the referee writes, 'He has very neat handwriting'. However, suppose further that the referee goes on to add, 'which aids engagement with his innovative philosophical ideas'. In this case, the implicature that the candidate is not suitable is cancelled, because an alternative explanation has been provided for the utterance, 'He has very neat handwriting', i.e., the

that both Grice's notion of 'saying' and his notion of 'implicating' incorporate his telling mechanism, as Neale observes:

> A necessary condition on conversational implicatures […] is that they are *intended.* [W]hat *U* implicates is part of what *U* means, and […] what *U* means is determined by *U*'s communicative intentions. A hearer may think that, by saying that *p*, *U* has conversationally implicated that *q* […] But if *U* did not intend the implication in question it will not count as a conversational implicature. Grice himself was explicit about it as far back as 'The Causal Theory of Perception' at p. 130 of the 1961 version. (1992: 528)

As such, the co-operative principle, along with the maxims that fall within it, represent the standards that hearers expect of speakers when in engaged in a maximally effective exchange of information, whether their 'contribution' is one of saying or implicating. As Grice says:

> [W]hat is implicated is what it is required that one assume a speaker to think in order to preserve the assumption that that he is observing the Cooperative Principle (and perhaps some conversational maxims as well), *if not at the level of what is said, at least at the level of what is implicated.* (1989: 86 – emphasis added)

Now, although I think Grice's framework of communication is best placed to help undertake our ambitious task, there are three immediate problems that arise for Grice's framework of communication. I call them the 'inaccurateness problem', the 'incompleteness problem', and the 'obscureness problem'.[92] I offer solutions to each of these problems in §4.3.1, §4.3.2, and

---

referee has provided information as to why the candidate's handwriting is relevant. (Note, this creative example of the way in which the referee may cancel their implicature is borrowed from Bourne & Caddick Bourne (2018: 154).) For more on Grice's account of cancelling, see his (1975: 39).

[92] I appreciate there may be other potential problems with Grice's framework of communication, but I do not have space to address all potential problems here.

§4.3.3., respectively. But let us take a moment to acquaint ourselves with these three problems, in anticipation of the solutions to come.

4.2.1. **The Inaccurateness Problem**

Although the maxims of quality are formulated in light of the telling mechanism, they seem to capture inaccurately the standards that hearers expect of speakers when engaged in a communicative exchange the purpose and direction of which is to inform each other of the facts. The super maxim does not require truth of the speaker but requires merely that the speaker only *try* to contribute the truth, which is consistent with the contribution being false. The first specific maxim requires merely that the speaker only *not disbelieve* what they say, which is consistent with the speaker being agnostic about what they say. And the second specific maxim requires that the speaker only say what they have *adequate evidence* for, which is rather vague. Gazdar (1979), dissatisfied with Grice's formulation of the maxims of quality, writes:

> Any attempt to formalise this maxim as it stands runs into three sets of problems. Those connected with the notion of "truth," those connected with the logic of belief, and those involved in the nature of "adequate evidence." Note, however, that these three sets of problems are just those that crop up in the philosophical debate over the status of knowledge and the possibility of equating knowledge with justified true belief. (47)

Not wishing to engage with the debate over the status of knowledge and the possibility of equating knowledge with justified true belief, Gazdar simply stipulates that Grice's maxims of quality should be reformulated into the following single maxim:

> The Maxim of Quality: Say only that which you know. (48)

Benton (2016), however, argues that Grice makes important connections between the maxims of quality and knowledge, which reveals that knowledge, for Grice, is in fact the

core requirement. He attempts to show this by drawing on the following example from Grice:

> *A*: Where does *C* live?
>
> *B*: Somewhere in the south of France.
>
> This infringement of the first maxim of Quantity [make your contribution as informative as is required (for the current purposes of the exchange)] can be explained only by the supposition that *B* is aware that to be more informative would be to say something that infringed the second maxim of Quality, "Don't say what you lack adequate evidence for," so *B* implicates that he does not *know* in which town *C* lives. (Grice 1989: 32–33 – emphasis Benton's)

According to Benton, *B*'s response is best explained by *B*'s sensitivity to a requirement of knowledge. That is, he thinks it is because *B* does not know exactly where *C* lives, and realises this, that *C* opts for the hedged claim. He maintains that by invoking *B*'s lack-of-knowledge implicature, Grice really suggests that knowledge, and not some weaker epistemic position, is intimately connected to whether a speaker is in a position to make a contribution.

Benton raises a further clash between the maxims of quantity and the maxims of quality in another attempt to show that knowledge is the core requirement of the maxims of quality. He invites us to consider cases in which a question is asked directly about whether a certain proposition, *p*, is true. As he observes, if an individual feels that they cannot answer directly in such circumstances, they will often reply, 'I do not know'. Benton then appeals to a series of unpublished papers by Grice in which he regards the 'I do not know' response as conforming to the maxims of quality at the expense of violating the maxims of quantity, due to providing too little information.[93] Benton argues that if the maxims of quality do not

[93] Grice Papers, 1947-1989: carton 1, folder 23, after page 6.

demand knowledge, it is hard to see how the 'I do not know' response could be perceived to be a clash between the requirements of the maxims of quantity and the maxims of quality. Moreover, he argues that if we took the maxims of quality to be lax enough to permit non-known contributions, then the 'I do not know' reply would seem to flout the maxims of relation as well, since it would be irrelevant and unhelpful for an individual to reply that they do not know. But, as he points out, when prompted by questions the answer to which an individual does not know (or takes themselves not to know), we do not judge the 'I do not know' response as irrelevant or unhelpful. Benton expands on this point by drawing on Turri (2010), who maintains that prompting questions, 'Do you know whether *p*?' and 'is *p*?', are taken to be practically interchangeable in everyday conversation. According to Benton, this suggests something important about the relation of the maxims of quality to our standard expectations of what counts as an appropriate response to a prompting question. Specifically, he claims that if the maxims of quality enjoin us to contribute only what we know, we would expect asking whether *p*, and asking whether someone knows that *p*, to be practically interchangeable, since individuals answer each question appropriately, and felicitously, with a contribution that *p* just in case they know that *p*. Relatedly, he argues that weaker prompts, 'Do you have good reason to think *p*?' and 'Do you have any idea whether *p*?', are not practically interchangeable with 'Is *p*?', which, he thinks, suggests that the maxims of quality demand something stronger than merely having good reasons or some evidence to believe that *p.* He concludes once more that knowledge, and not some weaker epistemic standard, is intimately connected to whether a speaker is in a position to contribute.

Benton's arguments go some way towards providing evidence that the most plausible interpretation of Grice's maxims of quality is that they require knowledge, even if Grice does not explicitly endorse this himself. However, such evidence is hardly definitive, nor does it take us much further than what Grice thought. What we need is a principled explanation – a rationale – for why the maxims of quality are best understood in terms of requiring knowledge. A solution to the inaccurateness problem is still needed.

4.2.2. **The Incompleteness Problem**

According to Grice, in cases of telling, the speaker exercises intentions to influence the hearer's beliefs in a certain way, namely, the speaker intends that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*. But how exactly does the hearer's recognition of the speaker's intention influence the hearer's beliefs? Grice writes:

> [T]he intended effect must be something which in some sense is within the control of the audience, or that in some sense of 'reason' the recognition of intention behind [the utterance] is for the audience a reason and not merely a cause. (1957: 385)

As Ross (1986) explains, Grice wants us to think of the hearer as having been furnished with a 'reason' in something more like the sense in which an individual has been provided with a motive for adopting a certain course of action. However, since Grice does not elaborate on this point, it is not easy to see what exactly he has in mind. Ross agrees with Grice that a hearer's response to what is communicated reflects a sensitivity to the speaker's intentions, but he suggests that the real motivation a hearer possesses for believing what is told stems from the speaker assuming responsibility for the truth of what is told.[94] Unfortunately, it is not clear how exactly this solves the problem. To see why, consider the following from Lackey (2008):

> [S]uppose that I walk up to an airplane mechanic and say, "I will take responsibility for the proper functioning of this aircraft." Given that I know nothing about the mechanics of aircrafts, my words would rightly be regarded as hollow even if said with the utmost sincerity and goodwill. (228)

[94] Moran (2005; 2018) famously attempts to solve the incompleteness problem along these lines.

Lackey's point is not how an assumption of responsibility for the truth is connected to the truth, but whether it has a proper connection to the truth at all. If, as Ross suggests, the motivation a hearer possesses for believing what is told stems directly from the speaker's responsibility for the truth of what is told, then when a speaker takes responsibility for the truth of what is told, even when a speaker communicates something false and the hearer knows this, the hearer nevertheless possesses a motivation for believing what they are told. Not only is this an unattractive result, but it also raises another question: in virtue of what precisely does the speaker assume responsibility for the truth of what is told? A solution to the incompleteness problem is still needed.

4.2.3. **The Obscureness Problem**

For convenience, Grice proposes that the conversational maxims which fall within the cooperative principle hold an equal status among one another. However, he acknowledges that the maxims of quality may have a special status:

> It is obvious that the observance of some of these maxims is a matter of less urgency than is the observance of others; a man who has expressed himself with undue prolixity would, in general, be open to milder comment than would a man who has something he believes to be false. Indeed, it might be felt that the importance of at least the first [specific] maxim of Quality is such that it should not be included in a scheme of the kind I am constructing; other maxims come into operation only on the assumption that this maxim of Quality is satisfied. While this may be correct, so far as the generation of implicatures is concerned it seems to play a role not totally different from the other maxims, and it will be convenient, for the present at least, to treat it as a member of the list of maxims. (Grice 1989: 27)

Grice's claim is not that the first specific maxim of quality is merely more important than the other conversational maxims, but that it has a distinct status among them. What, though, is

the status of the first specific maxim of quality? Grice does not offer an answer. However, Benton attempts to provide us with one.[95]

Benton agrees with Grice that the cooperative principle is a general principle, whereas the first specific maxim of quality enjoys a distinctive status among the other conversational maxims that fall within it.[96] For him, the fact that Grice claims that the first specific maxim of quality might best be given a class by itself, and that the other conversational maxims operate only on the assumption that it is satisfied, is the first step toward realising that it is the 'epistemic demands' of the first specific maxim of quality in particular that explain why it takes priority over the demands of other conversational maxims. Moreover, he observes that there is a prima facie irrelevance of the first specific maxim of quality with respect to speech acts such as questions and imperatives. Taking these two points together, he claims that the first specific maxim of quality is applicable to only a certain type of speech act, namely, assertion.[97] He draws on the following passage from Grice to bolster this idea:

> The maxims do not seem to be coordinate. The [first specific] maxim of Quality, enjoining the provision of contributions which are genuine rather than spurious (truthful rather than mendacious), does not seem to be just one among a number of recipes for producing contributions; it seems rather to spell out the difference between something's being, and (strictly speaking) failing to be, any kind of

[95] Benton, like many others, seems to treat Grice's work in the philosophy of language as comprising two unrelated theories (i.e., the 'theory of meaning' and the 'theory of conversation'), since he talks about the co-operative principle and conversational maxims with no mention of the telling mechanism.

[96] Benton does not always make it entirely clear whether he is talking about the super-maxim of quality, the first or second specific maxim of quality, or some combination. I do my best to clarify this for the reader.

[97] Reformulated for clarity from Benton (2016: 686– 687). Note, Benton thinks that the first specific maxim of quality is applicable only to the speech act of assertion. However, he does not offer an account of assertion.

> contribution at all. False information is not an inferior kind of information; it just is not information. (1989: 371)[98]

Benton maintains that we need not take the last line of Grice's passage literally but that we can nonetheless appreciate Grice's key point, which, according to him, is not only that the first specific maxim of quality is applicable to only a certain type of speech act (i.e., assertion), but that when a speaker makes as if to adhere to the first specific maxim of quality, when they in fact do not, they are thereby not being cooperative.[99] Here, Benton introduces his own technical term, 'respect':

> [O]ne […] respects a maxim when exploiting it to conversationally implicate, and one also trivially respects it when engaging in other speech acts, such as asking a question'. (2016: 12)

His rationale for introducing the notion of 'respect' is due to Grice's remark that the other conversational maxims come into operation only on the assumption that the first specific maxim of quality is satisfied. According to Benton, Grice does not mean to rule out cases in which one conversationally implicates by exploiting a maxim, since working out implicatures depends on the other maxims being operative. With this set out, Benton formulates an argument to try and show that the status of the first specific maxim of quality is such that it is not dependent on the cooperative principle in the way that the other conversational maxims are:

---

[98] As far as I am aware, this is the only time in Grice's work where he says, 'the maxim of Quality'. It is not entirely clear what he is referring to, whether it is the super-maxim, one of the first or second specific maxims, or some combination of them. Thank you to Deirdre Wilson (personal correspondence) for bringing this to my attention.

[99] Note, it is not clear why we would not say the same about the any of the other conversational maxims.

Premise 1: If the first specific maxim of quality is not presumed (by the audience) to be respected (by the speaker), then the other maxims do not operate.

Premise 2: If the other maxims do not operate, the cooperative principle cannot be operative.

Conclusion: Unless the first specific maxim of quality is presumed (by the audience) to be respected (by the speaker), the cooperative principle cannot be operative.[100]

The problem with Benton's argument, however, is that it does not square with the conversational evidence. That is, inferences that have the marks of calculating implicatures are often made by hearers even when they recognise that the speaker does not satisfy (or 'respect') the first specific maxim of quality. Stokke (2016) demonstrates this with the following case:

Thelma and Louise

> Thelma attempts to deceive Louise about the fact that she is OK to drive. However, Louise knows that Thelma has been drinking (and thus knows that she is not OK to drive), but Thelma does not know that Louise knows this.
>
> Louise: Are you OK to drive?
> Thelma: I haven't been drinking. (450)

Louise takes Thelma as having implicated that she is OK to drive by means of exploiting the maxim of relation in uttering that she has not been drinking, yet Louise knows that Thelma is nonetheless violating the first specific maxim of quality. (Of course, Thelma thinks that

[100] Note, this seems to show that 'respecting' the first specific maxim of quality is conditional on the cooperative principle being employed, not that it is independent of it. Indeed, I develop a similar idea in §4.3.3.

Louise will take her as satisfying (or 'respecting') the first specific maxim of quality to generate the implicature in question.) Since the implicature is still calculated by Louise, it is not the case that if the maxim of quality is not presumed (by the audience) to be respected (by the speaker), then the other maxims do not operate – the maxim of relation is plainly in operation, after all. Therefore, premise 1 of Benton's argument is false. A solution to the obscureness problem is still needed.

4.3. **Positive Proposal**

With my epistemological commitments and Grice's framework of communication set out, I am almost ready to construct my novel theoretical elaboration of our candidate explicatum, (AE). In order to do this, however, I must construct an account of the conversational dynamics which underpin communicative exchanges generally. I will then this use to construct an account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as the position is traditionally construed – in light of which I can construct my theoretical elaboration of our candidate explicatum. Given the ambitiousness of this task, it is important I take things slowly. Let us start, then, by fleshing out just the initial details of our candidate explicatum, (AE):

> (AE) Essentially, assertion is a communicative act which involves a speaker saying that *p* with the intention that the hearer acquire a belief that *p*.

As I see it, Grice's notion of saying is best construed as delineating the speech act of assertion as I understand it.[101] That is, (AE) should be fleshed out in the following way. Essentially, assertion is when a speaker says that *p* (in the Austinian sense (see §2.1)) with

---

[101] As we saw in §2.1.1 and §2.5, I am not the first to use Grice's work on telling to provide an account of assertion.

the intention that the hearer acquire a belief that *p* on the basis of the hearer recognising their intention that the hearer acquire a belief that *p*. On my view, then, comprehending a speaker as making an assertion involves assuming that they satisfy Grice's telling mechanism. Furthermore, assuming that the speaker satisfies the telling mechanism involves employing the expectation that the speaker adheres to certain standards which hold during our communicative exchanges generally, within which fall the conversational maxims specific to maximally effective exchanges of information.[102] Importantly, though, since a maximally effective exchange of information just is a communicative exchange the purpose or direction of which is to inform each other of the facts by means of the telling mechanism, I think such communicative exchanges are best construed as delineating communicative exchanges the purpose or direction of which is to transmit knowledge specifically. Hopefully, so far so good.

Having fleshed out the initial details of our candidate explicatum, I will now move on to construct my account of the conversational dynamics which underpin communicative exchanges generally. I will then use this to construct my account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as it is traditionally construed – in light of which I can finalise the construction of my theoretical elaboration of our candidate explicatum. Doing all of this, though, involves understanding Grice's communicative standards (i.e., the cooperative principle and the maxims that fall within it) in a certain way.

---

[102] Although I employ Grice's framework to articulate my account of assertion, my aim here is not to provide a piece of scholarly work on Grice's thought. Rather, I interpret and develop Grice's framework in a way that makes sense to me, which means that I may sometimes diverge from Grice's thought. As such, my account of assertion might be best understood as 'neo-Gricean' rather than 'Gricean'.

Although Grice's communicative standards might typically be understood in terms of rules (or norms), I do not understand them in this way. Instead, I think they are best understood in terms of the mutual expectations that speakers and hearers have about their contributions. That is, I think the cooperative principle encapsulates the expectations that speakers and hearers expect each other to expect each other to have about their contributions when engaged in communicative exchanges generally, and I think the cooperative principle and conversational maxims together encapsulate the expectations that speakers and hearers expect each other to expect each other to have about their contributions when engaged in communicative exchanges the purpose or direction of which is to transmit knowledge specifically.[103] That is not all, though. Crucially, I think it is the mutual expectations that speakers and hearers have about their contributions which determine Grice's communicative standards, rather than the other way around. Although our projects are significantly different, Goldberg (2018) endorses a view which is kindred in spirit, namely, a view according to which social and epistemic standards generally are determined by the social and epistemic expectations we have of each other, rather than the other way around. He captures this idea neatly in what he calls his 'order-of-explanation thesis'. I recast it here, in my own terms, for my own purposes:

> It is natural to think that our communicative standards are themselves independent of our mutual expectations of each other as communicators. That is, it is natural to think that such communicative standards rationalise or justify our mutual expectations of each other as communicators. According to this picture, we mutually expect individuals to conform to these communicative standards, precisely because these are the communicative standards. We might cast this picture in terms of an order-of-explanation thesis of the sort that is familiar from the Euthyphro dialogue: mutually we expect as much from individuals as communicators *because these are the communicative standards*. I propose to depart from this way of thinking on this point,

---

[103] Thank you to Craig Bourne and Emily Caddick Bourne (personal correspondence) for many helpful conversations about mutual expectations in relation to Grice's framework of communication.

> reversing the order of explanation: our communicative standards are what they are *because we have mutual expectations of individuals as communicators*. In other words, I do not regard the communicative standards themselves as independent of our mutual expectations, rather our mutual expectations themselves determine our communicative standards.[104]

It will be helpful to explain the source of my inspiration for understanding Grice's standards of communication in the way that I do. The source of my inspiration comes from Lewis's (1969; 1975) work on convention. According to Lewis, conventions are regularities in behaviour within a population, established by the mutual expectations and derived from the common interests of the individuals within said population. Consider the following example:

> [S]everal of us are driving on the same winding two-lane roads. It matters little to anyone whether he drives in the left lane or the right lane, provided the others do likewise. But if some drive in the left lane and some in the right, everyone is in danger of collision. So each must choose whether to drive in the left lane or in the right lane, according to his expectations about the others: to drive in the left lane if most or all of the others do, to drive in the right lane if most or all of the others do. (1969: 6)

Lewis calls situations such as this 'coordination problems'. The idea is that when two or more individuals are faced with a coordination problem each must choose one action from

---

[104] To emphasise, the idea here is this. Providing an account of communicative standards in terms of rules (or norms), especially those with an epistemological dimension, is somewhat mysterious. And so, an account that explains them in terms of something which is clear and plain, i.e., expectations, is highly desirable. Indeed, it constitutes significant philosophical progress. Note, Goldberg's original formulation of his order-of-explanation thesis may be found in his (2018: 145–189). Importantly, though, unlike Goldberg, who uses 'expectation' in the normative sense (i.e., expectations about what *ought* to happen), I use 'expectation' in the non-normative sense (i.e., expectations about what *will* happen).

two or more alternative actions, where the outcomes they want to produce or prevent are determined jointly by the actions of all the individuals involved, and where the outcome of any action an individual might choose depends on their expectations about what the others will do. Successful combinations of the individuals' chosen actions are called 'equilibria', i.e., combinations in which each individual has done as well as they can, given the actions of the other individuals. In the coordination problem above, the common interest is in being able to travel as safely as possible, i.e., without colliding. Thus, the two possible equilibrium combinations are the regularity that everyone drives in the left lane, or the regularity that everyone drives in the right lane. Notably, on Lewis's picture, explicit agreement is not necessary to generate conventions. That is, coordination problems may be solved via agreement, salience, or precedent. However, according to Lewis, whatever route individuals take in finding a solution to a coordination problem, they are most likely to succeed through the agency of a system of suitably concordant mutual expectations. Lewis's inspiration here comes from Schelling's (1960) work on coordination:

> What is necessary is to coordinate predictions, to read the same message in the common situation, to identify the one course of action that their expectations of each other can converge on. They must "*mutually recognise*" some unique signal that coordinates their expectations of each other. Most situations […] provide some clue for coordinating behaviour, some focal point for *each person's expectation of what the other expects him to expect to be expected to do*. (Schelling 1960: 54–57 – emphasis added)

In a nutshell, Lewis's account of convention is as follows. A regularity *R* is a convention in a population *P* if and only if, within *P*, everyone conforms to *R*, everyone expects that others conform to *R*, the presence of the expectation that others conform to *R* serves as the explanation for why everyone conforms to *R* themselves, and everyone prefers general conformity to *R* rather than conformity by all but one. Vitally, for *R* to be a convention in *P* it cannot be the only possible regularity. That is, there must be at least one alternative regularity, *R**, which could have perpetuated itself instead of *R* to realise the common interest in the specific aim/outcome the population has in that particular context. This last

feature of Lewis's account is designed to capture the characteristic arbitrariness of conventions.[105]

Lewis famously uses his account of convention to explain what it is for it to be a convention among a population to use some particular language over some alternative language, e.g., what it is for it to be a convention among a population to speak English rather than Mandarin. He claims that if we look for the fundamental difference in verbal behaviour between members of two distinct linguistic communities, we find something which is arbitrary but perpetuates itself because of a common interest in communication, which derives from a common interest in benefitting from, and in preserving, our ability to control each other's beliefs and actions. His analysis is as follows:

> A language £ is used by a population *P* if and only if there prevails in *P* a convention of truthfulness and trust in £, sustained by an interest in communication.[106]

While constructing his analysis, however, Lewis poses an objection against it:

> *Objection*: Truthfulness and trust cannot be a convention. What could be the alternative to uniform truthfulness – uniform untruthfulness, perhaps? But it seems that if such untruthfulness were not intended to deceive, and did not deceive, then it too would be truthfulness. (Lewis 1975: 29)

He then provides the following reply to this self-imposed objection:

> *Reply*: The convention is not the regularity of truthfulness and trust *simpliciter*. It is the regularity of truthfulness and trust in some particular language £. Its alternatives

---

[105] Admittedly, Lewis's picture is slightly more complicated than this. For example, he has further conditions which draw on the notion of common knowledge. However, this simplified account will suffice for our purposes.

[106] Simplified from Lewis (1975: 9).

> are possible regularities of truthfulness and trust in other languages […] There is a different regularity that we may call a regularity of truthfulness and trust *simpliciter*. That is the regularity of being truthful and trusting in whichever language is used by one's fellows. This regularity neither is a convention nor depends on convention. If any language whatever is used by a population *P*, then a regularity […] of truthfulness and trust simpliciter prevails in *P*. (29–30)

Now, the source of my inspiration for understanding Grice's communicative standards the way that I do is not Lewis's explanation of what it is for it to be a convention within a population to use some particular language over some alternative language, but rather what he calls a 'regularity of truthfulness and trust *simpliciter*'. As can be seen from the passage above, he thinks that the regularity of truthfulness and trust simpliciter is not a convention, because it is not arbitrary. That is, there is no alternative regularity which could have perpetuated itself instead of truthfulness and trust simpliciter to realise the common interest in communicating true information, true information that is relevant to which actions will produce desired results, which itself derives from a common interest in benefitting from, and in preserving, our ability to control each other's beliefs and actions. As such, the regularity of truthfulness and trust simplicter is best understood as falling under Lewis's related though distinct concept of a 'demythologised social contract'.[107] Lewis maintains that a so-called 'demythologized social contract' is similar to but not the same as a convention, because there is no alternative regularity in behaviour which can realise the common interest in the specific aim/outcome the population has in that particular context. And so, it is not the case that everyone prefers general conformity rather than conformity by all but

---

[107] See Lewis (1969: 88–96; 1975: 30–31). Note, Hobbes (1651) provides us with what is arguably the most developed picture of a 'mythologized' social contract. He begins by inviting us to consider a primitive unstructured order, which he calls the 'state of nature'. Here, individuals have unlimited natural freedom to further their own selfish interests, including the freedom to harm anyone who threatens their own self-preservation. This thus leads to generalised conflict and instability. He maintains therefore that it is rational for individuals to relinquish some of their unlimited natural freedom, so that they may enjoy the benefits that social structure and civil rights provide. Doing so results in a 'mythologized' social contract.

one, as with convention. Instead, everyone prefers general conformity to a certain contextually definite state of general nonconformity, i.e., a so-called 'demythologized state of nature'. As such, a so-called 'demythologized social contract' is underpinned by a mixture of selfish interests, altruistic interests, and interests derived from moral obligation, where usually all are present in strength, but where usually any one is enough to sustain it.[108] That is, individuals do not merely have interests which fit a model of selfish interest, but also have non-selfish interests due to the way they have been socialised (at home and/or at school) to be helpful, get along, or at least not go out of their way to make things difficult for others.[109]

Having sketched out the source of my inspiration for understanding Grice's communicative standards the way that I do, I can now construct my account of the conversational dynamics which underpin communicative exchanges generally. I will then use this to construct my account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-

---

[108] See Lewis (1975: 30).

[109] Note, Lewis does not offer up or draw upon any particular account of trust, but merely remarks that the appropriate response to truthfulness for a hearer, i.e., trusting, involves believing the speaker. However, there is an account of trust on the market which shores up neatly with Lewis's picture and which shores up neatly with my theoretical elaboration of our candidate explicatum as well, namely, Hawley's (2012) commitment account of trust. On Hawley's account, to trust someone to do something is to believe that they have a commitment to doing it, and to rely upon them to meet that commitment. Accordingly, to distrust someone to do something is to believe that they have a commitment to doing it, and yet not rely upon them to meet that commitment. To make the account plausible, she uses a very broad notion of commitment. As she says, commitments on her account can be implicit or explicit, weighty or trivial, conferred by roles and external circumstances, default or acquired, welcome or unwelcome. In particular, though, she maintains that mutual expectations as well as conventions give rise to commitment, unless we take steps to disown these. Although I will not say too much more about trust, this is the account I am sympathetic to and which readers should have in mind as I work through my theoretical elaboration of our candidate explicatum.

reductionism as it is traditionally construed – in light of which I can construct my theoretical elaboration of our candidate explicatum. Here goes.

As I see it, Gricean cooperation derives from a common interest in communication generally, which itself derives from a common interest in benefitting from and preserving the ability to control one another's beliefs and actions. It is preferable to a state in which individuals cannot communicate with each other and thus benefit from, and preserve, the ability to control one another's beliefs and actions. And so, it is underpinned by a mixture of selfish interests, altruistic interests, as well as interests derived from moral obligation, where usually all are present in strength, but where usually any one is enough to sustain it. The standard of cooperation itself, however, is determined by the mutual expectation that individuals will do all and only that which is required in order to realise the common interest in communication. Specifically, it is determined by the mutual expectation that individuals will make their contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the communicative exchange in which they are engaged. As such, the mutual expectation of cooperation does not merely determine the standard of cooperation, it also constitutes the conditions for communication generally. Importantly, it is the presence of the mutual expectation of cooperation, engendered whenever an individual makes a contribution, which serves as the explanation for why individuals conform to the standard of cooperation *by default*. Why? Not merely because there is future gain to be had from conforming (i.e., selfish interests), but because they have been socialised (at home and/or at school) to be helpful, get along, or at least not go out of their way to make things difficult for others (i.e., altruistic interests and interests derived from moral obligation). Of course, if an individual defects, they are liable to experience personal hardship for undermining their own integrity and/or receive sanctions. Thus, the benefit that an individual might enjoy for being uncooperative must be significant enough to make doing so rational, in combination with the likelihood of feeling like they have let

others (e.g., the hearer) as well as themselves down and/or receiving sanctions.[110] However, this is not to say that individuals conform to the standard of cooperation out of fear or concern for experiencing personal hardship for undermining their own integrity and/or receiving sanctions necessarily. They might conform simply because the thought of defecting does not cross their minds. Given that the presence of the mutual expectation of cooperation, engendered whenever an individual makes a contribution, serves as the explanation for why individuals conform to the standard of cooperation *by default*, the mutual expectation of cooperation itself is a mutual *default* expectation. That is, individuals are not merely psychologically disposed to expect that contributors will do all and only that which is required in order to realise the common interest in communication, they are entitled (albeit defeasibly) to expect that contributors will do all and only that which is required in order to realise the common interest in communication *by default*.[111] To put it in Lewis's epistemological terms, although the possibility that an individual is uncooperative cannot be eliminated, it can be (properly) ignored. Finally, Gricean cooperation is not a convention, because there is no alternative mutual default expectation which can be satisfied, and thus no alternative standard which can be met, in order to realise the common interest in communication.

This is my account of the conversational dynamics which underpin communicative exchanges generally. I can now use this to construct my account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically, which, in turn, will provide us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as the position is traditionally construed. As I see it, communicative exchanges the purpose or direction of which is to transmit knowledge specifically are an extension of – and indeed conditional on

---

[110] Simion (2020), a proponent of the constitutive rule (or norm) account of assertion, argues for anti-reductionism as the position is traditionally construed using Hobbes's 'mythologized' social contract (mentioned in fn.107), part of which involves employing a similar sort of schema to the one I mention here.

[111] This of course should not be read as involving a vicious circularity, but rather read as involving a legitimate form of bootstrapping.

– the conversational dynamics which underpin communicative exchanges generally, and so they *embody* the conversational dynamics which underpin communicative exchanges generally, as follows.

Truthfulness derives from a common interest in transmitting knowledge specifically, which itself derives from a common interest in benefitting from and preserving the ability to communicate true information about the environment to each other reliably, true information that is relevant to which actions will produce desired results. It is preferable to a state in which individuals cannot engage in transmitting knowledge and thus communicate true information about the environment to each other reliably, true information that is relevant to which actions will produce desired results. And so, it is underpinned by a mixture of selfish interests, altruistic interests, as well as interests derived from moral obligation, where usually all are present in strength, but where usually any one is enough to sustain it. The standard of truthfulness itself, however, is determined by the mutual expectation that individuals will do all and only that which is required in order to realise the common interest in transmitting knowledge. Specifically, it is determined by the mutual expectation that individuals will tell that *p* in virtue of the fact that the reason why the tell that *p* is that they know that *p*, a relevant corollary of which is the mutual expectation that what they tell, *p*, is true. As such, the mutual expectation of truthfulness does not merely determine the standard of truthfulness, it also constitutes the conditions for knowledge transmission. Importantly, it is the presence of the mutual expectation of truthfulness, engendered whenever an individual makes a contribution which involves the telling mechanism specifically, which serves as the explanation for why individuals conform to the standard of truthfulness *by default*. Why? Not merely because there is future gain to be had from conforming (i.e., selfish interests), but because individuals have been socialised to be helpful, get along, or at least not go out of their way to make things difficult for others (i.e., altruistic interests and interests derived from moral obligation). Of course, if an individual defects, they are liable to experience personal hardship for undermining their own integrity and/or receive sanctions. Thus, the benefit that an individual might enjoy for being untruthful must be significant enough to make doing so rational, in combination the likelihood of feeling like they have let others (e.g., the hearer) as well as themselves down

and/or receiving sanctions. However, this is not to say that individuals conform to the standard of truthfulness out of fear or concern for experiencing personal hardship for undermining their own integrity and/or receiving sanctions necessarily. They might conform simply because the thought of defecting does not cross their minds. Given that the presence of the mutual expectation of truthfulness, engendered whenever an individual makes a contribution which involves the telling mechanism specifically, serves as the explanation for why individuals conform to the standard of truthfulness *by default*, the mutual expectation of truthfulness itself is a mutual *default* expectation.[112] That is, individuals are not merely psychologically disposed to expect that contributors will do all and only that which is required in order to realise the common interest in knowledge transmission, they are entitled (albeit defeasibly) to expect that contributors will do all and only that which is required in order to realise the common interest in knowledge transmission *by default*. To put it in Lewis's epistemological terms, although the possibility that an individual is untruthful cannot be eliminated, it can be (properly) ignored. Finally, truthfulness is not a convention, because there is no alternative mutual default expectation which can be satisfied, and thus no alternative standard which can be met, to realise the common interest in transmitting knowledge.[113]

Given that communicative exchanges the purpose or direction of which is to transmit knowledge specifically are an extension of – and indeed conditional on – the conversational dynamics which underpin communicative exchanges generally, we can see now that the nature of the interpersonal relation at play between speakers and hearers in cases of knowledge transmission is such that the default position for speakers *is* to assert the truth.

---

[112] Again, this of course should not be read as involving a vicious circularity, but rather read as involving a legitimate form of bootstrapping.

[113] I appreciate that this might be seen as gratuitously repetitious, but I think that repetition is a helpful device here for seeing clearly the sense in which communicative exchanges the purpose or direction of which is to transmit knowledge specifically are an extension of – and indeed conditional on – the conversational dynamics which underpin communicative exchanges generally, and thus the sense in which they *embody* the conversational dynamics which underpin communicative exchanges generally.

As such, hearers *do* enjoy a defeasible entitlement to believe speakers, and thus what is asserted as true, by default. Thus, it *is not* the case that on top of comprehending the speaker as having made an assertion the hearer need be aware of additional positive reasons for believing what is asserted.

We now have a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as it is traditionally construed – in light of which I can finalise the construction of my theoretical elaboration of our candidate explicatum. I think this will be achieved most advantageously, however, by working through the three problems for Grice's framework of communication, mentioned above: the inaccurateness problem, the incompleteness problem, and the obscureness problem. I start with the inaccurateness problem.

4.3.1. **The Inaccurateness Problem Revisited**

As we saw in §4.2.1, Benton's arguments go some way towards providing evidence that the most plausible interpretation of Grice's maxims of quality is that they require knowledge, even if Grice does not explicitly endorse this himself. However, such evidence is hardly definitive, nor does it take us much further than what Grice thought. What we need is a principled explanation – a rationale – for why the maxims of quality are best understood in terms of requiring knowledge. Fortunately, we are now in a position to offer one.

Although the co-operative principle and maxim categories hold during our communicative exchanges in general, the conversational maxims themselves do not, since Grice formulates these with a specific type of communicative exchange in mind, namely, a maximally effective exchange of information. A maximally effective exchange of information just is a communicative exchange the purpose or direction of which is to inform each other of the facts by means of the telling mechanism. On my interpretation, such communicative exchanges are best construed as delineating communicative exchanges the purpose or direction of which is to transmit knowledge specifically. Accordingly, when a speaker makes an assertion that *p*, they engender the mutual default expectation that what they assert, *p*, is

true, which is a corollary of an ancillary mutual default expectation they engender when they make an assertion that *p*, namely, that they assert that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*. This explains why knowledge is the requirement of the maxims of quality. Why? Because if the specific ancillary mutual default expectation required anything weaker than knowledge, e.g., belief, not only would this undermine the possibility of transmitting knowledge, but it would also undermine the idea that there is a mutual default expectation that what the speaker asserts, *p*, is true – a mere belief that *p* might well be false, after all. We might say, then, that when a speaker makes an assertion that *p* there is a 'super' mutual default expectation that what the speaker asserts, *p*, is true, which is a corollary of a 'specific' mutual default expectation that the speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*.[114] I reformulate Grice's maxims of quality, accordingly:

---

[114] Sperber & Wilson (2002) argue that truth cannot be the requirement for assertion because speakers make assertions that are only approximated and so not strictly true. For example, when someone asks, 'What is the time?' or 'How tall are you?', speakers typically assert answers like, 'It is 5pm' and 'I am 6ft', when in fact it is a few seconds or minutes before or after 5pm, and when in fact they are a few millimetres or centimetres taller or shorter than 6ft. However, by drawing on Schoubye & Stokke's (2016) account of 'what is said', I believe we can solve this apparent problem. As Schoubye & Stokke argue, what is said by an utterance of a sentence in a context is determined by the answers to 'questions under discussion' (or QUDs). On this view, contexts are assumed to contain QUDs as formal objects, and sentences are analysed as true or false relative to the QUD addressed in that context. However, what a sentence can be used to express, i.e., what its potential truth conditions are, is constrained by a compositionally derived minimal content. So, while a single sentence has a wide variety of potential truth conditions depending on the contextually relevant QUD being addressed, this set of potential truth conditions is constrained by the meanings of its constituents and its syntax. Crucially, though, Schoubye & Stokke do not identify minimal compositional meaning with the truth conditions of a sentence relative to the context. Therefore, when a speaker says, 'It is 5pm' or 'I am 6ft', this minimal compositional meaning alone does not determine what is said, but rather the QUD, say, 'Approximately (give or take a minute or so), what is the time?' or 'Approximately (give or a take a centimetre or so), how tall are you?', along with the minimal compositional meaning, determines what is said. In these two cases, then, what is said is, 'Approximately (give or take a minute or so), it is 5pm' and 'Approximately (give or a take a centimetre or so), I am 6ft tall'. If so, contra Sperber & Wilson, what is asserted (because what is said) is strictly true.

Maxims of quality

The super-maxim:

- Assert that *p* only if *p* is true.

The specific maxim:

- Assert that *p* only if you know that *p*.[115]

Since the communicative standard of truthfulness is determined by the mutual default expectations encapsulated in the maxims of quality, a speaker is truthful in what they assert when and only when they satisfy the specific mutual default expectation and thus the super mutual default expectation.

4.3.2. **The Incompleteness Problem Revisited**

As we saw in §4.2.2, according to Grice, in cases of telling, the speaker exercises intentions to influence the hearer's beliefs in a certain way, namely, the speaker intends that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*. However, it seems there is something missing in Grice's account, specifically, the mere recognition of a speaker's intention does not seem to provide an epistemic basis for the hearer to believe anything. Grice wants us to think of the hearer as having been furnished with a 'reason' in something more like the sense in which an individual has been provided with a motive for adopting a certain course of action. Why, though, would the hearer's recognition of the speaker's intentions influence the hearer's beliefs in this way? The following passage from Grice offers a clue:

---

[115] I have simplified the specific maxim here, for the mutual default expectation is that the speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*.

> On my account, it will not be true that when I [assert] that *p*, I conversationally implicate that I believe that *p*; for to suppose that I believe that *p* […] is just to suppose that I am observing the first maxim of Quality on this occasion. I think that this consequence is intuitively acceptable; it is not a natural use of language to describe one who has [asserted] that *p* as having, for example, "implied," "indicated," or "suggested" that he believes that *p*; the natural thing to say is that he has expressed […] the belief that *p*. He has of course committed himself, in a certain way, to its being the case that he believes that *p*, and while this commitment is not a special case of [asserting] that he believes that *p*, it is bound up, in a special way, with [asserting] that *p*. (Grice 1989: 42)

What Grice's passage seems to suggest is that the speech act of assertion and the maxims of quality have a special relationship. Recall, however, that on my interpretation the maxims of quality are understood in terms of the mutual default expectations between the speaker and the hearer. Specifically, they are understood in terms the specific mutual default expectation that the speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary. Moreover, on my interpretation, said mutual default expectations are engendered whenever a speaker makes a contribution which involves the telling mechanism specifically. Strictly speaking, then, it is the telling mechanism and said mutual default expectations which have a special relationship. As such, the maxims of quality should be understood as having a more general form, as follows:

> Maxims of quality
>
> The super-maxim:
>
> - Tell that *p* only if *p* is true.

The specific maxim:

- Tell that *p* only if you know that *p*.[116]

What precisely is the nature of the special relationship between the telling mechanism and said mutual default expectations, though? I submit that it is an *exclusive symbiotic* relationship. Let me explain what I mean by this. When a speaker intends that a hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*, the very recognition of this intention is comprised of the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. Furthermore, if a speaker makes a contribution which does not involve the intention that a hearer acquire a belief on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief, then the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tells that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary, simply is not engendered. In other words, the maxims of quality are contribution-specific, namely, they are specific to contributions which involve the telling mechanism.[117] Therefore, the speech act of assertion

---

[116] This is as it should be, given that that the maxims of quality are applicable not only to the speech act of assertion, but to implicature as well – another realisation of the telling mechanism. (I discuss this in §4.3.5). Note, I have again simplified the specific maxim here, for the mutual default expectation is that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*.

[117] Notice that this idea suggests that different contributions have their own unique contribution-specific maxims, all of which fall under the maxim *category* of quality. For example, Schaffer (2006) suggests that the speech act of asking questions is perhaps beholden to certain standards, such as: 'ask only when you do not know'. We might say, then, that comprehending a speaker as asking a question involves the mutual default expectation that the speaker asks whether *p* in virtue of the fact that the reason why they ask whether *p* is that they do not know that *p*. If so, we would say that this maxim is also a maxim of quality, but a maxim of quality engendered when asking questions specifically, and so distinct from the maxims of quality engendered when telling. Nonetheless,

and the maxims of quality have a special relationship only because the speech act of assertion is a *realisation* of the telling mechanism. With this in place, the solution to the incompleteness problem presents itself.

When a speaker makes an assertion, they freely and publicly publish their intention that the hearer acquire a belief on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief. In doing so, they freely and publicly engender the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. Why? Because due to the exclusive symbiotic relationship between the telling mechanism and the maxims of quality, i.e., the mutual default expectations that they encapsulate, these mutual default expectations are comprised of the mutual recognition of the speaker's reflexive intention, and the mutual recognition of the speaker's reflexive intention is comprised of said mutual default expectations. This thus affirms the significance of the self-referential nature of the telling mechanism: the speaker's reflexive intention provides the hearer with a 'reason' to believe what is told, because it is mutually understood between the speaker and hearer what the speaker is doing in exercising it, namely, inviting the hearer to trust them. And unless there is some relevant defeater present, the hearer is entitled (albeit defeasibly) to accept the speaker's invitation by default. We might say, then, that the telling mechanism is imbued with epistemic value in virtue of the mutual default expectations of which it is comprised, and thus, as a realisation of the telling mechanism, the speech act of assertion is endowed with epistemic import.

Notice that this solution to the incompleteness problem makes good sense of the somewhat mysterious comments Grice makes regarding attitude expression and commitment. That is, by appealing to the mutual default expectations of which the telling mechanism is

---

the maxim of quality for asking questions, just like the maxims of quality for telling, would fall under the maxim *category* of quality, as all contribution-specific maxims would.

comprised, we can see the sense in which a speaker, when they make an assertion that *p*, freely and publicly confers a commitment upon themselves to it being the case that they assert that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, and thus that what they assert, *p*, is true.[118] As such, we can see the sense in which a speaker, when they make an assertion, not only expresses or manifests a belief in what they assert, but indeed represents themselves as having the knowledge in question – even if they neither know nor believe what they assert. Lastly, we can see the sense in which a speaker, when they make an assertion, freely and publicly incurs a responsibility for the belief that the hearer acquires (if they acquire it). Importantly, though, it is neither the commitment the speaker confers upon themselves, nor their expression of belief or having represented themselves as knowledgeable, nor the responsibility they incur in making an assertion which explains why the hearer's recognition of the speaker's reflexive intention influences the hearer's beliefs. Rather, it is the mutual default expectations of which the speaker's reflexive intention is comprised which explain this.

Notice further that this solution also explains challenges to and retractions of assertions. That is, by appealing to the mutual default expectations of which the telling mechanism is comprised, we can see why a speaker, when they make an assertion, might be met with challenges such as, 'That is not true', 'How do you know that?' or 'You do not believe that!'. If the hearer doubts that the speaker satisfies the super mutual default expectation that what the speaker asserts, *p*, is true and/or the specific mutual default expectation that the speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, they might issue a challenge of this sort. Accordingly, we can see that a retraction of an assertion would be called for whenever a speaker asserts under conditions in which they do not satisfy said mutual default expectations. Making such a retraction, in turn, would make it clear to the hearer that the speaker is not available to shoulder the epistemic burden for the proposition in question.

---

[118] We can get a sense here of the way in which Hawley's account of trust and my theoretical elaboration of our candidate explicatum complement one another.

4.3.3. **The Obscureness Problem Revisited**

As we saw in §4.2.3, for convenience, Grice proposes that the conversational maxims which fall within the cooperative principle hold an equal status among one another. However, he acknowledges that the maxims of quality may have a special status. Specifically, he claims that the other maxims come into operation only on the assumption that the first specific maxim of quality is satisfied. Benton's attempt to vindicate this claim fails. What, then, should we make of this?

On my interpretation, communication generally is conditional on the cooperative principle, i.e., the mutual default expectation that the speaker makes a contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the communicative exchange in which they are engaged. As such, when a speaker makes a contribution which involves the telling mechanism specifically, comprehending the speaker involves assuming that they satisfy the telling mechanism, which itself involves employing the cooperative principle as well as the maxims that fall within it. We might say, then, that the cooperative principle and the maxims that fall within it *work in the service of* contribution comprehension, contributions which involve the telling mechanism specifically. Moreover, on my interpretation, the maxims of quality are understood as a particularised extension of the cooperative principle, exclusively for contributions which involve the telling mechanism specifically, and thus exclusive to communicative exchanges the purpose or direction of which is to transmit knowledge specifically. Indeed, this is the sense in which the maxims of quality have a special status.

Notice, however, that contribution comprehension is compatible with the possibility that a speaker quietly and unostentatiously violates the maxims of quality, even if the hearer is aware of the speaker's artfulness. Why? Because although the cooperative principle, and thus the maxims that fall within it, i.e., the mutual default expectations that they encapsulate, are defeasible, it does not follow from this that a speaker's artfulness, and indeed the hearer's awareness of a speaker's artfulness, results in all-out abandonment of said mutual default expectations. Such a situation is of course one in which the hearer is not

entitled to expect that the speaker tells that *p* in virtue of the fact that the reason why the speaker tells that *p* is that the speaker knows that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. Nevertheless, the hearer expects the speaker to expect them (the hearer) to expect that they (the speaker) tell that *p* in virtue of the fact that the reason why the speaker tells that *p* is that the speaker knows that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. And the speaker expects the hearer to expect them (the speaker) to expect that they (the speaker) tell that *p* in virtue of the fact that the reason why the speaker tells that *p* is that the speaker knows that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. In other words, even though the speaker quietly and unostentatiously violates the maxims of quality, and even if the hearer is aware of the speaker's artfulness, said mutual default expectations are employed in such a way that enables the hearer to comprehend the speaker's contribution, regardless of the fact that the hearer ultimately is not entitled to believe what the speaker tells them, and likely does not believe what the speaker tells them.[119] Thus, Grice and Benton are mistaken in thinking that any of the maxims of quality must be satisfied in order for the other conversational maxims, and indeed the cooperative principle, to come into operation.

4.3.4. **Speaker Sincerity and Speaker Reliability**

According to my theoretical elaboration of our candidate explicatum, the specific mutual default expectation that a speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p* – of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary – may be conceptually split into mutual default

[119] If, however, the situation was such that both the speaker and the hearer were mutually aware that the speaker at least seemingly fails to satisfy the cooperative principle, and thus the maxims that fall within it, i.e., the mutual default expectations that they encapsulate, contribution comprehension would not be possible. I discuss this in relation to Moore's paradox (i.e., item (11) on the substantive in §2.5) in the next chapter (§5.2.2).

expectations regarding the speaker's sincerity and the speaker's reliability, i.e., regarding the speaker being both honest in what they tell and actually having the knowledge in question.

Following Stokke (2018), I think that speaker sincerity is best understood in terms of assent.[120] Assent is the mental action of an individual occurrently putting forward a proposition as true in their mind – conscious mental affirmation.[121] Assent is different from, but intimately related to, the state of belief. Ordinarily, an individual who assents to *p* holds the corresponding belief that *p*. However, it is possible for assent and belief to come apart, i.e., for an individual to assent to *p* but not hold the corresponding belief that *p*. Moreover, it is possible for this to happen and for it not to be transparent to the individual in question, or anyone else. The reason why an individual might assent to *p* but not hold the corresponding belief that *p* is due to non-rational factors, such as self-deception, prejudice, and phobias.[122] Accordingly, I think that speaker reliability is best understood in terms of the speaker holding the corresponding belief to the proposition they assent to and that belief amounting to knowledge. Ordinarily, when a speaker tells that *p*, they are both sincere and reliable. However, it is possible for a speaker's sincerity and reliability to come apart, i.e., for a speaker to assent to *p* but not hold the corresponding belief that *p* (due to non-rational factors), or, if they do hold the corresponding belief that *p*, for that belief that *p* not to amount to knowledge that *p*. Moreover, it is possible for this to happen and for it not to be transparent to the individual in question, or anyone else. The reason why an individual might assent to *p* and hold the corresponding belief that *p* but not have knowledge that *p* is because said belief that *p* does not satisfy the conditions to ascribe knowledge that *p* to that individual in that particular context, as sketched out in §4.1. To be clear, then, the mutual

---

[120] Admittedly, sincerity is traditionally thought of as a matter of a speaker telling what they believe. However, due to various (I think, convincing) counterexamples in the philosophy of language literature, this traditional view is not so plausible. See Chan and Kahane (2011) and Stokke (2018) for a comprehensive exposition of the sincerity debate.

[121] As Stokke observes, 'assent' and 'judgement' are often used interchangeably. However, like him, I stick with the former term.

[122] Again, see Cassam (2010)

default expectation that a speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p* (of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary) may be conceptually split into the mutual default expectation that they tell that *p* in virtue of the fact that the reason why tell that *p* is that they:

(i) assent to *p*
(ii) hold the corresponding belief that *p*,
(iii) and said belief that *p* amounts to knowledge that *p*.

Therefore, a speaker is insincere when they do not aim to satisfy the specific mutual default expectation that they tell that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary. That is, a speaker is insincere when they tell that *p* but it is not the case that the reason why they tell that *p* is that they assent to *p*. Accordingly, a speaker, on the assumption that they are sincere, is unreliable when they do aim to satisfy the specific mutual default expectation that a speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary, but unbeknownst to them they fail to do so. That is, a speaker, on the assumption that they are sincere, is unreliable when they tell that *p* in virtue of the fact that the reason why they tell that *p* is that they assent to *p*, but unbeknownst to them they do not hold the corresponding belief that *p*, or they do hold the corresponding belief that *p* but unbeknownst to them said belief that *p* does not amount to knowledge that *p*.

At this juncture, I would like to flag an uncertainty of mine. As just mentioned, according to my theoretical elaboration of our candidate explicatum, the specific mutual default expectation that a speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p* – of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary – may be conceptually split into mutual default expectations regarding the speaker's sincerity and the speaker's reliability, i.e., regarding the

speaker being both honest in what they tell and actually having the knowledge in question. However, although what I have proposed so far in my theoretical elaboration of our candidate explicatum seems to vindicate the mutual default expectation that speakers are sincere, does any element of it vindicate the mutual default expectation that speakers are reliable? For instance, a speaker, due to the way they have been socialised to be helpful, get along, or at least not go out of their way to make things difficult for others, may aim to satisfy the specific mutual default expectation that they tell that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, of which the super mutual default expectation that what they tell, *p*, is true is a corollary. However, unbeknownst to the speaker (and perhaps unbeknownst to the hearer they address as well), they may fail to do so. Why? Either because they tell that *p* in virtue of the fact that the reason why they tell that *p* is that they assent to *p*, but unbeknownst to them they do not hold the corresponding belief that *p*, or because they tell that *p* in virtue of the fact that the reason why they tell that *p* is that they assent to *p* and hold the corresponding belief that *p*, but said belief that *p* unbeknownst to them does not amount to knowledge that *p*. Both cases are ones in which the speaker is sincere but unreliable. It might seem, then, that regardless of what I have proposed so far, the possibility of a speaker's unreliability is such that it flies in the face of the idea that there is a specific mutual default expectation that a speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary. In other words, what I am wondering is this: although my theoretical elaboration explains why the possibility that a speaker is lying can be (properly) ignored, strictly speaking does it explain why the possibility that a speaker is irrational or incompetent can be (properly) ignored as well? If not, then my theoretical elaboration of our candidate explicatum is incomplete and thus needs supplementing.[123]

---

[123] As I say, I am uncertain whether my theoretical elaboration of our candidate explicatum is incomplete in the sense I have described. For instance, it might be thought that what I have said so far already implies that speakers are sufficiently vigilant in selecting what they choose to tell and what they choose not to tell, such that, in ordinary contexts, what is told is told sincerely and reliably. After all, what is clear from my theoretical elaboration of our candidate explicatum is that making an assertion involves both the speaker and the hearer

On the assumption that my theoretical elaboration of our candidate explicatum is incomplete in the sense that I have described, I suggest tentatively that a promising way to supplement it might be found in Davidson's (1973; 1984; 1991; 2001) work on interpretation.[124]

According to Davidson, in order to interpret the language of another speaker, an interpreter must construct a manual that enables the assignment of truth conditions to any sentence of the language under interpretation, where the manual is constructed by means of discerning patterns in the relations between the uttered sentences the speaker holds true and the objects and events in the speaker's environment that prompt them to assent to the truth of the sentences they utter. He exemplifies his account of interpretation by focusing on what he calls 'radical interpretation', which is a matter of interpreting the speaker of a hitherto unknown language without relying on any prior knowledge either of the speaker's beliefs or the meanings of the speaker's utterances. What such cases draw out is that an interpreter cannot assign meanings to a speaker's utterances without knowing what the speaker believes, and an interpreter cannot identify beliefs without knowing what the speaker's utterances mean. Davidson responds to this challenge, however, by suggesting that we must assign meanings to speakers' utterances and identify the relevant beliefs simultaneously. He maintains that this can be achieved via the application of what he calls the 'principle of

---

understanding what it is the speaker is doing and what it is that is expected of the speaker in doing it. If this is the case, then we might already have a straightforward explanation not only for why the possibility that a speaker is, say, lying can be (properly) ignored, but also why the possibility that a speaker is, say, irrational or incompetent can be (properly) ignored. Moreover, it is not obvious that this problem (if it is a problem) is specific to my view. That is, even though the possibility that speakers are insincere when engaged in the act of telling is (properly) ignorable by hearers, if the possibility that speakers are irrational or incompetent when engaged in the act of telling is a live possibility for hearers (i.e., not (properly) ignorable), then this potentially undermines any anti-reductionist position as it is traditionally construed, specifically, conditions (AR2) and (AR3).

[124] I am by no means the first to draw upon Davidson's account of interpretation to defend anti-reductionism as it is traditionally construed. For alternative applications see Coady (1992) and Stevenson (1993).

charity', a principle which he holds is constitutive of interpretation generally, not just radical interpretation:

> Principle of charity: when interpreting a speaker's linguistic behaviour, optimise agreement between oneself and the speaker in such a way as to maximize the number of true beliefs attributed to the speaker.[125]

For Davidson, the principle of charity encapsulates the ineliminable (albeit defeasible) presuppositions that must be employed in order to interpret speakers' utterances. More specifically, an interpreter, on the presupposition that they themselves are rational and respond to their environment in such a way as to acquire true beliefs, must presuppose that the speaker under interpretation is rational and responds to their environment in such a way as to acquire true beliefs as well. Said presuppositions enable the interpreter to ascribe true beliefs to the speaker (albeit defeasibly), which, in turn, enables the interpreter to assign meaning to the speaker's utterances (albeit defeasibly). Davidson's justification for the principle of charity rests on the claim that without the presuppositions encapsulated within it, interpretation is not possible. Why? Because, he claims, for interpretation to be possible, the interpreter must be able to accept or reject certain interpretations of a speaker's utterances, and this is only intelligible against a background of massive agreement between the interpreter and the speaker, which the presuppositions encapsulated within the principle of charity provide. In other words, the principle of charity is the bedrock of the possibility of interpretation altogether. He writes:

> If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything. (1973: 137)

---

[125] Davidson's principle of charity admits of various formulations and so is difficult to articulate precisely, but this will suffice for our purposes.

Since Davidson's account of interpretation – especially his principle of charity on which his entire account of interpretation is built – enjoys some support within the philosophical community, it is not unreasonable to suggest tentatively that it might offer a promising way to supplement the apparent incompleteness of my theoretical elaboration of our candidate explicatum.[126] What, then, is my tentative suggestion? To explain this, I first need to make a revision to Davidson's principle of charity. As we have seen, the principle of charity does not merely encapsulate the presuppositions that interpreters are psychologically disposed to employ when they interpret speakers' linguistic behaviour. Rather, it encapsulates the ineliminable (albeit defeasible) presuppositions that interpreters must employ in order to interpret speakers' linguistic behaviour at all. Moreover, as we have seen, what interpreters presuppose essentially is that the speaker under interpretation is rational and responds to the environment in such a way as to acquire true beliefs, just like the interpreter. This, as we have seen, enables the interpreter to ascribe true beliefs to the speaker (albeit defeasibly), which, in turn, enables the interpreter to assign meaning to the speaker's utterances (albeit defeasibly). We might ask, though, what is it for an individual to respond to an environment in such a way as to acquire true beliefs? Put slightly differently, we might ask, how does an individual respond to an environment in such a way as to acquire true beliefs instead of falling into error? My answer to this question, as we have seen in §4.1, is reliabilist in nature, namely, by bringing it about that the processes by which they acquire beliefs are truth conducive. Indeed, as we also saw in §4.1, I think that knowledge in general is the product of forming true beliefs in this way. As such, I think that what interpreters ascribe to speakers under interpretation (albeit defeasibly) is not merely true belief, but knowledge. I appreciate that there is a lot more to say in filling out and defending such a revision to the principle of charity, but to ward off any immediate worries, it is worth noting that an almost identical revision to the principle of charity is already filled out and defended by Williamson (2000; 2007). He writes:

---

[126] For endorsements/developments, see Grandy (1973), Lewis (1974), Williamson (2000; 2007), and Goldberg (2015).

> The appropriate principle of charity will give high marks to interpretations on which speakers tend to assert what they *know*, rather than to those on which they tend to assert what is true, or even what is reasonable for them to believe. (2000: 267 – emphasis added)

With this revision in place, I am ready to offer up my tentative suggestion: when a speaker makes a contribution which involves the telling mechanism specifically, the hearer must employ the ineliminable (albeit defeasible) presuppositions that the speaker is rational and knowledgeable, and they must employ said ineliminable (albeit defeasible) presuppositions in combination with the cooperative principle and the maxims that fall within it, i.e., the mutual default expectations that they encapsulate, in order to interpret the speaker's utterance and thus comprehend the speaker as satisfying the telling mechanism.

Supplemented in this way, my theoretical elaboration of our candidate explicatum does not merely vindicate the mutual default expectation that speakers tell that *p* in virtue of the fact that the reason why they tell that *p* is that they assent to *p*, it also vindicates the mutual default expectation that speakers tell that *p* in virtue of the fact that the reason why they tell that *p* is that they assent to *p*, hold the corresponding belief that *p*, and that said belief that *p* amounts to knowledge that *p*. In other words, supplemented in this way, my theoretical elaboration of our candidate explicatum not only explains why the possibility that a speaker is lying can be (properly) ignored, it also explains why the possibility that a speaker is irrational or incompetent can be (properly) ignored. And so, we can now recast Lewis's passage (from §4.1) on the overall reliability of perception in terms of the overall reliability of assertion:

> What non-circular argument can be given for our reliance on assertion? Assertions depend on the speaker, and what we believe about the speaker depends in turn on their assertions. Each dependence covers a wide and varied range of alternatives. Of course, it is possible for the speaker to be insincere, irrational, or incompetent. Such

> possibilities never can be eliminated. But they can be ignored. And if they are properly ignored – as they mostly are – then assertions give us knowledge.

To reiterate, I am uncertain whether my theoretical elaboration of our candidate explicatum even is incomplete in the sense I have described. Moreover, on the assumption that it is, what I have provided is only a tentative suggestion for how it might be supplemented. Still, what I aim to have shown here is that I know what I have got an argument for and where I propose to look for support for the extra step, if I need it.

4.3.5. **Assertion vs. Implicature**

On my view, just like Grice's, the notion of telling cuts across assertion and implicature.[127] As such, when a speaker generates an implicature, they freely and publicly publish their intention that the hearer acquire a belief on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief. In doing so, they freely and publicly engender the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. By appealing to the mutual default expectations of which the telling mechanism is comprised, we can see the sense in which a speaker, when they implicate that *p*, freely and publicly confers a commitment upon themselves to it being the case that they implicate that *p* in virtue of the fact that the reason why they implicate that *p* is that they know that *p*, and thus that what they implicate, *p*, is true. As such, we can see the sense in which a speaker, when they generate an implicature, not only expresses or manifests a belief in what they implicate, but indeed represents themselves as having the knowledge in question – even if they neither know nor believe what they implicate. Plus, we can see the sense in which a speaker, when they generate an implicature, freely and publicly incurs a responsibility for the belief that the hearer acquires

---

[127] Of course, Grice would hold that that the notion of telling cuts across *saying* and implicature, rather than *assertion* and implicature, since he does not talk in terms of assertion, as I do.

(if they acquire it). Accordingly, on my view, implicatures, just like assertions, are apt to transmit knowledge.

Fricker (2012), however, disagrees. According to her, in making an assertion, a speaker incurs full responsibility for the truth of what is communicated, whereas, in generating an implicature, they do not. Essentially, her argument takes the following form:

Premise 1:	If a speaker's communicative act is governed by the standard of knowledge, then the speaker incurs full responsibility for the truth of *p* by performing that communicative act.
Premise 2:	A speaker does not incur full responsibility for the truth of *p* by generating an implicature that *p*.
Conclusion:	Implicature is not a communicative act which is governed by the standard of knowledge.

Fricker's argument stands or falls on the strength of premise 2, for which she provides a set of interlocking supporting arguments. Let us take them in turn. First, she makes the observation that, in cases of assertion, the speaker's reflexively intended meaning is made explicit by virtue of the symbols used (because the speaker's reflexively intended meaning is essentially the same as what is said (in the Austinian sense (see §2.1)), which the 'surrounding linguistic context' suffices to fix and allow the recovery of the speaker's reflexive intentions. However, in cases of implicature, the speaker's reflexively intended meaning is not made explicit by virtue of the symbols used (because the speaker's reflexively intended meaning is something over and above what is said, but instead depends on the 'shared knowledge context' and Grice's conversational maxims to recover the speaker's reflexive intentions. Based on this observation, she argues that, in cases of implicature, unlike cases of assertion, it is unclear to hearers what proposition is communicated, if there even is a single proposition communicated. Indeed, she maintains that what is communicated via implicature is more like a fuzzy set of propositions. Second, she argues that implicature thus demands far more from the hearer than assertion does to work out correctly what the speaker's reflexively intended meaning is. Third, she argues

that, in cases of implicature, unlike in cases of assertion, the speaker thus enjoys plausible deniability regarding what is communicated. She concludes, therefore, that, in cases of implicature, unlike cases of assertion, a speaker cannot plausibly be understood as incurring full responsibility for the truth of what they communicate.

However, I do not find Fricker's interlocking arguments in support of premise 2 compelling. Let me explain. Although Fricker's initial observations about assertion and implicature seem to be acceptable, it is not obvious that her first argument follows from these observations. Why? Two reasons. First, as Grice makes clear, for an implicature to count as an implicature, it must be capable of being worked out. He writes:

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature. (Grice 1975: 31)[128]

Thus, Fricker's argument that, in cases of implicature, unlike cases of assertion, it is unclear to hearers what proposition is communicated is simply at odds with Grice's stipulation of what implicature is, and so she is not giving implicature a fair footing from the off. Second, and as a way of bolstering her first argument, Fricker maintains that what is communicated via implicature is more like a fuzzy set of propositions. The problem with this, however, is that it is applicable to assertion as well. That is, we successfully communicate all the time by asserting such things as, 'He is a bit aggressive', which is of course compatible with a fuzzy

---

[128] Grice (1975) provides a general pattern for the working-out of implicatures, as follows:

> He has said that *q*; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that *p*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *p* is required; he has done nothing to stop me thinking that *p*; he intends me to think, or is at least willing to allow me to think, that *p*; and so he has implicated that *p*. (31)

set of specific propositions about how aggressive the individual is. What a speaker communicates (whether by means of implicature or assertion) is almost never without some degree of vagueness, after all. As for Fricker's argument that implicature demands far more from the hearer than assertion does to work out correctly what the speaker's reflexively intended meaning is, I struggle to see what the relevance of this is. That is, even if implicature does demand more of the hearer than assertion does in order to work out correctly what the speaker's reflexively intended meaning is, and even if there is, for this reason, more responsibility on the hearer to work it out correctly, it is hard to see what this has to do with the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. In other words, it seems quite plain that there are two different types of responsibility at play in cases of implicature, just as there are in cases of assertion, namely, the speaker's responsibility regarding the truth of what is communicated and the hearer's responsibility regarding working out correctly what is communicated. As I see it, even if it is true that that implicature demands far more from the hearer than assertion does to work out correctly what the speaker's reflexively intended meaning is, this is orthogonal to the speaker's responsibility regarding the truth of what is communicated. Lastly, Fricker's argument that, in cases of implicature, unlike in cases of assertion, the speaker enjoys plausible deniability regarding what is communicated simply does not have the force which she takes it to have. Why? Two reasons. First, both implicature and assertion suffer from plausible deniability. To see this in the case of assertion specifically, consider the following example from Peet (2015):

Special Craft Lager

> Suppose it is the start of a new year and we have organised a party for the new graduate students. We have a variety of beers on offer, but there are some special craft lagers I want for myself (even though they were brought for the guests). I have stored most of the beer in the fridge, but I have put the craft lagers outside. Sally, one of the new students, arrives and asks where the beer is, so I tell her 'all the beer is in

> the fridge'. Later on you find the craft lagers outside and ask me why I told Sally that all the beer was in the fridge. In response to this challenge I might attempt to construct a story along the following lines: I had heard that Sally was a vegan, and I am aware that craft lagers often contain animal products. So when I said 'all the beer is in the fridge' I didn't mean every beer we had purchased for the party, I meant every beer which was safe for Sally, as a vegan, to drink. (33)

This observation is of course not intended to spell trouble for assertion in the same way that Fricker tries to spell trouble for implicature. Instead, it is intended to level the playing field by showing that there is no difference in principle between implicature and assertion on the matter of plausible deniability, thus taking the sting out of Fricker's argument. Admittedly, the plausible deniability for assertion is comparatively lower in degree than it is for implicature. However, the plausible deniability of implicature is also often exaggerated. And this is my second point. Recall Grice's famous case of the referee who writes a reference for a candidate for an academic position but only utters, 'He has very neat handwriting'. Given that the referee is supposed to focus on academic expertise, the information that they provide seems to be irrelevant, as well uninformative. In order to reconcile the speaker's utterance with the maxims of relation and the maxims of quantity, the hearer takes the referee as being as relevant and as informative as it is possible to be within the remit of the task, which is to provide positive information about the candidate. So, the hearer calculates that the referee has used their utterance of 'He has very neat handwriting' to implicate that the candidate is not suitable for the position. Suppose, though, that the recipient of the reference comes to know the candidate personally through other avenues and (quite unprofessionally) later challenges the referee for having told them (and the rest of the academic board) that the candidate in question was not suitable. Suppose further that the referee responds to the recipient's challenge by saying, 'I didn't tell you that! I said that they have very neat handwriting. I certainly did not intend for you to believe that they are not a suitable candidate'. The referee's response would of course be frustrating for the reference recipient because the referee would have changed the topic of conversation from one about the (il)legitimacy of what they told the reference recipient (and the rest of the academic board) to one about the technicalities of communication. More importantly than that,

though, the referee's denial would not be plausible. Why? Because the assumption here is that it is mutually understood between the referee and the referee recipient (and the rest of the academic board) that, when engaged in writing and providing the reference, the question under discussion would have been, 'Is the candidate suitable for the position?'. As such, for the referee to deny that they intended the reference recipient (and the rest of the academic board) to believe that the candidate in question was not suitable for the position by uttering, 'He has very neat handwriting', implies that the referee did not understand what the question under discussion was at the time, which is tantamount to renouncing any understanding of the type of communicative exchange they were engaged in at the time. Alternatively, it implies that the referee had a perfectly good understanding of the question under discussion at the time, and thus the type of communicative exchange they were engaged in at the time, but nevertheless decided to trigger the formal process of providing a reference to an institution just to pontificate irrelevantly about the candidate's handwriting. Neither of these options is plausible.[129] In sum, I do not find Fricker's interlocking arguments in support of premise 2 compelling and thus I submit that her argument overall is unsound.

4.4. **Summation**

In the previous chapter (§3.4), I said that knowledge transmission is the processes by which a speaker's knowledge that *p* becomes a hearer's knowledge that *p* (i.e., the hearer's knowledge that *p* is grounded upon the speaker's knowledge that *p*), in virtue of the fact that the reason why the speaker asserted that *p* is that the speaker knows that *p*, and whereby the content so asserted, *p*, comes to be known by the hearer by way of the hearer believing the

---

[129] The notion of a 'question under discussion' need not be formalised to understand my meaning here. If a formalisation were to be sought, though, Roberts (1996) would be my first port of call, with the idea being that what is implicated by an utterance of a sentence is determined (at least in part) by the answers to 'questions under discussion' (or QUDs), much like Schoubye & Stokke (2016) argue that what is said by an utterance of a sentence in a context is determined by the answers to 'questions under discussion' (or QUDs). (See fn.114.) This would provide a more theoretically robust way of understanding what is implicated by a speaker in a certain context, and thus the plausible deniability said speaker has with regard to the implicatures they generate.

speaker themselves. Moreover, I said that to fully understand knowledge transmission we must understand the nature of the interpersonal relation at play between speakers and hearers in such cases. Furthermore, I said that to understand this we must understand the conversational dynamics which underpin communicative exchanges generally. In this chapter (§4.3), using Grice's framework of communication, I provided an account the conversational dynamics which underpin communicative exchanges generally. This enabled me to provide an account of the nature of the interpersonal relation at play between speakers and hearers in cases of knowledge transmission. And this, in turn, provided us with a fully developed view of knowledge transmission – a view of knowledge transmission which satisfies (AR1), (AR2), and (AR3) of the anti-reductionist position as it is traditionally construed. In light of this fully developed view of knowledge transmission, I constructed my theoretical elaboration of our candidate explicatum Moreover, in the course of constructing my theoretical elaboration of our candidate explicatum, I solved three immediate problems for Grice's framework of communication as well as addressed items (1) through (10) on the substantive list (from §2.5). Plus, there is no immediate reason to think that my theoretical elaboration of our candidate explicatum will not afford us a vocabulary to help us explain any outstanding philosophically interesting phenomena peripheral in the study of assertion as well. Indeed, I addressed certain other philosophically interesting phenomena which are peripheral in the study of assertion in terms of it, namely, implicature and Davidson's theory of (radical) interpretation.[130] In the next chapter, I respond to some potential objections to my theoretical elaboration of our candidate explicatum. In addition, I address the final item, item (11), on the substantive list.

4.5. **Appendix C: a sneaky solution to sneaky intentions**

The formulation of the telling mechanism I use in my theoretical elaboration of our candidate explicatum is essentially the same as Grice's original formulation:

---

[130] Admittedly, if Davidson's theory of (radical) interpretation is to be understood in the way I present it in §4.3.4, then it is less peripheral (and indeed central) in the study of assertion than is often thought.

> "*U* meant something by uttering *x*" is (roughly) equivalent to "*U* intended the utterance of *x* to produce some response in an audience [e.g., the acquisition of belief] by means of the recognition of this intention". (Grice 1989: 220)

However, supposedly following a suggestion from Strawson, Grice replaces his original formulation with a second, 'iterated' formulation:[131]

> "By uttering *x*, *U* meant something" is true if and only if for some audience *A*, *U* uttered *x* intending:
>
> 1) *A* to produce some particular response *r* [e.g., the acquisition of belief]
> 2) *A* to recognize that *U* intends (1)
> 3) *A*'s recognition that *U* intends (1) to function, in part, as a reason for (1). (Grice 1989: 92)

As Neale (1992) observes, it is not entirely clear why Grice makes this replacement: he might see the iterated formulation as an elaboration on the original formulation (without involving any fundamental change), or he might see the iterated formulation as a modification of the original formulation (involving a fundamental change). I will not speculate on Grice's reasons for making this replacement. However, I see no fundamental difference between the two. As such, when I talk about the 'telling mechanism', I should be understood as referring to both formulations – they are essentially the same, as far as I am concerned, after all. Historically, Grice's telling mechanism has been thought to suffer from a debilitating problem introduced by Strawson (1964). Consider the following case:

[131] See Neale (1992) for a history of Grice's formulations.

Rat Infestation

> Suppose *A*, a friend of *U*, is about to buy a house, but *U* intends that *A* acquire a belief that the house is rat-infested, because *U* knows that it is. *U* lets rats loose in the house knowing that *A* is watching *U*. However, *U* knows that *A* does not know that *U* knows that *A* is watching *U* do this. Nonetheless, *U* knows that *A* will not take the presence of *U*'s rats as evidence that the house is rat-infested. But *U* knows, indeed *U* intends, that *A* will take the fact that *U* let rats loose in the house as evidence for thinking that *U* intends that *A* acquire a belief that the house is rat-infested. Furthermore, *U* knows that *A* has grounds for thinking that *U* would not intend that *A* acquire a belief that the house is rat-infested unless *U* has knowledge that it is.[132]

According to Strawson, in this case, conditions (1), (2), and (3) of the telling mechanism are satisfied, yet *U* cannot be accurately described as communicating with *A* in the sense that Grice seeks to elucidate, because *U*'s intentions are not 'wholly overt'. He writes:

> *A* will indeed take [*U*] to be trying to bring it about that *A* is aware of some fact; but he will not take [*U*] as trying, in the colloquial sense, to "let him know" something (or to "tell" him something). But unless [*U*] at least brings it about that *A* takes him [(*U*)] to be trying to let him [(*A*)] know something, he has not succeeded in communicating with *A*; and if, as in our example, he has not even tried to bring this about, then he has not even tried to communicate with *A* […] An essential feature of the intentions which make up the illocutionary complex is their overtness. They have, one might say, essential avowability. […] [T]he understanding of the force of an utterance in all cases involves recognizing what may be called broadly an audience-directed intention and recognizing it as wholly overt, as intended to be recognized. (447–459)

---

[132] Case adapted from Neale (1992), who adapts it from Strawson (1964) and Schiffer (1972).

Strawson is of course right to think that overtness in this sense is an essential feature of communication. Thus, given that the telling mechanism apparently violates this requirement, it seemingly allows for cases of communication in which all three conditions of the telling mechanism are satisfied but the utterer intends the audience to *believe* that they are not. Indeed, the utterer may intend that the audience believe that the utterer wants to produce some particular response in the audience (e.g., the acquisition of belief) for an entirely different reason. This has come to be known as the 'problem of sneaky intentions'. I believe, however, that a novel solution to the problem of sneaky intentions can be found by drawing on insights from an unrelated debate between Searle (1969) and Grice. Consider the following:

German Soldier

> Suppose that I am an American soldier in the Second World War [...] captured by Italian troops. And suppose [...] that I wish to get these troops to believe that I am a German soldier in order to get them to release me. What I would like to do is to tell them in German or Italian that I am a German soldier. But let us suppose I don't know enough German or Italian to do that. So I [...] attempt to put on a show of telling them that I am a German soldier by reciting those few bits of German I know, trusting that they don't know enough German to see through my plan. Let us suppose I know only one line of German which I remember from a poem I had to memorize in a high school German course. Therefore, I, a captured American, address my Italian captors with the following sentence: *Kennst du das Land wo die Zitronen blühen*? Now let us describe the situation in Gricean terms. I intend to produce a certain effect in them, namely the effect of believing that I am a German soldier, and I intend to produce this effect by means of their recognition of my intention. I intend that they should think that what I am trying to tell them is that I am a German soldier. But does it follow from this account that when I say, *Kennst du das Land … etc.*, what I mean is, "I am a German soldier"? (44)

Searle originally uses *German Soldier* as an objection to Grice's so-called 'intention-based semantics'. However, *German Soldier*, and importantly Grice's response to it, can be re-appropriated to shed light on the problem of sneaky intentions. Why? Because Searle's central claim is that *German Soldier* is a case in which conditions (1), (2), and (3) of the telling mechanism are satisfied, when, as Grice demonstrates, it is not. Grice responds to Searle, as follows:

> Is [*German Soldier*] […] a genuine counterexample? It seems to me that the imaginary situation is under-described […] The situation might be such that the only real chance that the Italian soldiers would, on hearing the American soldier speak his German line, suppose him to be a German officer, would be if they were to argue as follows: "He has just spoken in German (perhaps in an authoritative tone); we don't know any German, and we have no idea what he has been trying to tell us, but if he speaks German, then the most likely possibility is that he is a German officer – what other Germans would be in this part of the world?" If the situation was such that the Italians were likely to argue like that, and the American knew that to be so, then it would be difficult to avoid attributing to him the intention, when he spoke, that they should argue like that. As I recently remarked, one cannot in general intend that some result should be achieved, if one knows that there is no likelihood that it will be achieved. But if the American's intention was as just described, then he certainly would not, by my account, be meaning that he is a German officer; for though he would intend the Italians to believe him to be a German officer, he would not be intending them to believe this on the basis of their recognition of his intention. And it seems to me that though this is not how Searle wished the example to be taken, it would be much the most likely situation to have obtained. (Grice 1989: 101)

To be clear, the key point to draw out from Grice's response to Searle is that although *German Soldier* satisfies condition (1) of the telling mechanism, it fails to satisfy condition (2), and thus condition (3). That is, *German Soldier* is not best described as a case in which the American soldier intends that the Italian soldiers acquire a belief that he is a German soldier on the basis of *them recognising his intention that they acquire a belief that he is a German soldier.*

Rather, it is best described as a case in which the American soldier intends that the Italian soldiers acquire a belief that he is a German soldier on the basis of *an argument that they will likely construct for themselves*, as follows:

German Soldier Inference

Premise 1:	He has just spoken in German.
Premise 2:	If he speaks German, then he is a German soldier.
Conclusion:	He is a German soldier.

This observation is crucial to solving the problem of sneaky intentions, because, on reflection, we can see that *German Soldier* and *Rat Infestation* are in fact structurally analogous: *Rat Infestation* satisfies condition (1) of the telling mechanism, but it fails to satisfy condition (2), and thus condition (3). That is, *Rat Infestation* is not best described as a case in which *U* intends that *A* acquire a belief that the house is rat-infested on the basis of *A recognising U's intention that A acquire a belief that the house is rat-infested*. Rather, it is best described as a case in which *U* intends that *A* acquire a belief that the house is rat-infested on the basis of *an argument A will likely construct for themselves*, as follows:

Rat Infestation Inference

Premise 1:	*U* released rats into the house.
Premise 2:	If *U* released rats into the house, then *U* intends that I (*A*) acquire a belief that the house is rat-infested.
Premise 3:	If *U* intends that I (*A*) acquire a belief that the house is rat-infested, then the house is rat-infested.
Conclusion:	The house is rat-infested.

What this demonstrates is that Strawson does not provide a genuine counterexample to Grice's framework of communication. Why? Because he does not provide a case in which conditions (1), (2), and (3) of the telling mechanism are satisfied and in which the

communicators intentions are not wholly overt in the relevant sense. Notice, though, that *Rat Infestation*, unlike *German Soldier*, has the following feature: *A*'s inferences involve reference to *U*'s intentions (as can be seen in Premise 2 and Premise 3 of *Rat Infestation Inference*). My sneaky solution to the problem of sneaky intentions thus resides in the following observation: although *A*'s inferences involve reference to *U*'s intentions, it is a mistake to think that this alone supports the claim that *U* intends that *A* acquire a belief that the house is rat-infested on the basis of *A* recognising *U*'s intention that *A* acquire a belief that the house is rat-infested.[133] Strawson, however, mis-describes the case in just this way and thus makes precisely this mistake.

Note, we can vindicate this sneaky solution to the problem of sneaky intentions by reflecting on the *mutuality* involved in genuine cases of telling. Recall, when a speaker makes a contribution which involves the telling mechanism specifically, comprehending the speaker involves assuming that they satisfy the telling mechanism, which itself involves employing the cooperative principle, and thus the maxims that fall within it, i.e., the mutual default expectations that they encapsulate. That is, when a speaker makes a contribution which involves the telling mechanism specifically, they freely and publicly publish their intention that the hearer acquire a belief on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief. And in doing so, they freely and publicly engender the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, a corollary of which is the super mutual default expectation that what the speaker tells, *p*, is true. Why? Because due to the exclusive symbiotic relationship between the telling mechanism and the maxims of quality, i.e., the mutual default expectations that they encapsulate, said mutual default expectations comprise of the mutual recognition of the speaker's reflexive intention, and the mutual recognition of the speaker's reflexive intention is comprised of said mutual default expectations. As I see it, it is precisely this mutuality involved in genuine cases of telling

[133] Note, this seems to provide us with a kind of pattern for generating solutions to cases of this form, no matter how complex they may be.

which embodies Strawson's notion of 'essential avowabilty' and 'overtness'. However, as Strawson himself observes, *Rat Infestation* does not admit of such mutuality. Rather than thinking this spells trouble for Grice's notion of telling, though, it seems more appropriate to think it spells trouble for *Rat Infestation* being plausibly construed as a genuine counterexample to Grice's notion of telling.

5. Exploring the Exotica

In this chapter, I play the assertion counterexample game. I start by setting out my general approach to apparent borderline cases of and apparent counterexamples to my view of assertion, which essentially is to let my theoretical elaboration of our candidate explicatum, (AE), lead the way. I focus on three types of cases in particular, namely, so-called 'bald-faced lies', 'double bluffs', and 'resolute assertions' (introduced in §2.5). Each one of these types of cases apparently shows that it possible to make an assertion that *p* without the intention that the hearer acquire a belief that *p*, which would therefore mean that (AE) is too restrictive and thus implausible. However, by utilising my theoretical elaboration of our candidate explicatum, as well as drawing on other independently plausible theoretical considerations from the philosophy literature, I show that (AE) can be defended against these three types of cases. Moreover, in the course of doing this, I address the final item, item (11), on the substantive list (i.e., Moore's paradox). By the end of the chapter, I hope to have shown that playing the assertion counterexample game is not the daunting prospect that it is sometimes made out to be, but rather a useful philosophical tool.

5.1. **Arguing by Theory vs. Arguing by Counterexample**

At the beginning of this thesis (§2.5), I proposed the following candidate explicatum:

> (AE) Essentially, assertion is a communicative act which involves a speaker saying that *p* with the intention that the hearer acquire a belief that *p*.

In the previous chapter (§4.3), I provided a full theoretical elaboration of our candidate explicatum, as follows. Essentially, assertion is when a speaker says that *p* (in the Austinian sense (see §2.1)) with the intention that the hearer acquire a belief that *p* on the basis of the hearer recognising their intention that the hearer acquire a belief that *p*. As such, comprehending a speaker as making an assertion involves assuming that they satisfy the

telling mechanism. Furthermore, assuming that the speaker satisfies the telling mechanism involves employing the cooperative principle, and thus the maxims that fall within it, i.e., the mutual default expectations that they encapsulate. Specifically, it involves employing the specific mutual default expectation that the speaker tells that *p* in virtue of the fact that the reason why they tell that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker tells, *p*, is true is a corollary. Accordingly, the type of communicative exchange in which the speech act of assertion is performed are those the purpose or direction of which is to transmit knowledge specifically.

With this in hand, I am almost ready to play the assertion counterexample game, i.e., where I, having identified a subset of sayings in terms of the features detailed above, and essentialising and labelling the resulting act type 'assertion', am faced with counterexamples designed to show that it is possible to make an assertion which does not involve said essential features. Before I play, however, some comments are in order. As I said in §2.6, while some philosophers might hastily and wholeheartedly rely on arguments which turn on apparent borderline cases or apparent counterexamples as a way of guiding their inquiry, I am sceptical of this strategy. As I see it, trying to interpret and respond to apparent borderline cases or apparent counterexamples without first imposing some theoretical considerations on them is tantamount to philosophising inside a vacuum, where no traction can be gained, and which leads to a philosophical war of attrition in which each side fights using only their intuitions. As such, I think that arguments should instead turn on the theoretical elaboration of our candidate explicatum, which should occasion philosophers to scrutinise apparent borderline cases or apparent counterexamples more cautiously. That is, furnished with my theoretical elaboration of our candidate explicatum, we should feel comfortable to raise questions about the inherent plausibility of apparent borderline cases or apparent counterexamples, or at least be willing to entertain the possibility that they might be best understood in terms of something other than assertion, especially if not doing so might all too quickly cost us the theoretical elaboration of the concept and the fruit it bears. As I also said in §2.6, I appreciate that some might worry that this will lead to an unpalatable revisionism about certain apparent cases of assertion. However, I think such a worry is misplaced. Why? Because such a worry makes sense only if the apparent cases of assertion

under consideration are sufficiently well-established in the philosophy literature as cases of assertion in the first place in order to make revising them unpalatable. Yet, as we have seen, what 'assertion' denotes in the philosophy literature, apart from the paradigm case, is unclear.

In light of this, I wish to dismiss at the outset what might be seen as two apparent borderline cases of assertion and thus what might be seen as two potential counterexamples to my view: secret diary entries and soliloquies (see §2.2). Since speakers in such cases are typically not understood as directing their thoughts and feelings towards another person, they simply do not fit my conception of assertion. Moreover, I do not see any compelling reason why we must class them as cases of assertion. This should not be seen as a problem, though, not only because the objection of revisionism cannot be meaningfully levelled against me, but because I concede that they are nonetheless cases of testimony – indeed, cases of hearer-testimony. And, as we have seen, the term 'testimony' extends over different types of speech act, though it is unclear just how many different types of speech act it may extend over.[134] Thus, these cases already have their respective labels, and philosophers are already well positioned to theorise about them, without classing them as cases of assertion.[135] Some might flag up here that there are alternative conceptions of assertion in the philosophy literature on which cases of secret diary entries and soliloquies are classed as cases of assertion. In response, I invite them to consider the following three questions regarding any such alternative conception of assertion:

---

[134] To the best of my knowledge there is no comprehensive philosophical treatment of a secret diary entry or a soliloquy. Nonetheless, they have been subsumed under the moniker 'assertion' by philosophers of language and epistemologists of testimony alike.

[135] I should say, though, that I think it has been overstated how often hearers might acquire knowledge on the basis of finding secret diary entries and overhearing soliloquies. Secret diary entries are often the result of venting sessions, where the speaker expresses their feelings and emotions. Similarly, soliloquies are often performed by speakers who are venting, and thus expressing their feelings and emotions. In other words, in cases of secret diary entries and soliloquies, for the speaker, it is quite normal for truth to take a backseat.

- Can it, like the conception of assertion I have proposed, fruitfully be applied to almost all the philosophically interesting phenomena related to the speech act (i.e., all of items (1) through (11) on the substantive list in §2.5)?[136]
- Does it, like the conception of assertion I have proposed, at least seemingly afford us a vocabulary to help us explain the outstanding phenomena peripheral in the study of assertion?
- Is it, like the conception of assertion I have proposed, transparent with regard to the things that it postulates, the nature of those things, and the mechanisms involving them?

The invitation to consider such questions embodies the method of reflective equilibrium at play here, namely, to weigh up the cost/benefit of ruling out certain cases as cases of assertion against the cost/benefit of ruling in such cases as cases of assertion. As I see it, when it comes to ruling out secret diary entries and soliloquies as cases of assertion specifically, there is no cost apart from maybe some initial discomfort for some readers in reconciling their intuitions. However, the benefit of ruling out such cases as cases of assertion means that we get to keep hold of a comprehensive, fruitful, and transparent conception of assertion. Thus, I do not see my immediate dismissal of secret diary entries or soliloquies as cases of assertion as unreasonable, nor do I see it as showing that my conception of assertion is implausible in anyway.[137]

Essentially, I think more weight should be given to verdicts with the force of theory behind them. And since I think that my view is the best unified theory, I take my verdicts about

---

[136] I address the final item, item (11), on the substantive list (i.e., Moore's paradox) in §5.2.2.

[137] Admittedly, the implication here is that there is currently no alternative conception of assertion in the philosophy literature such that these questions can be answered affirmatively, let alone answered affirmatively and rule in secret diary entries and soliloquies as cases of assertion. Although this is what I think, I am not able to gather together all the existing conceptions of assertion in the philosophy literature in one place and show that this is the case and why. As such, the best I can do is invite the reader to consider such questions themselves and come to their own conclusions.

what counts as an assertion to be better for that reason. As such, I am making a suggestion about how I would categorise cases which, on my view, do not turn out to be assertions but concerning which there is at least some reason to think that they are assertions. That is, I am not trying to convince anyone that my re-categorisations must be correct but rather to show that I have a reasoned and defensible response to apparent counterexamples. Of course, a problem might arise if someone rejects the theory that I endorse, but that will not be because apparent counterexamples create a problem for my view while not creating a problem for others. In other words, my take on apparent counterexamples seems to be at least as good as any alternative. Therefore, objectors will have to do more than just assert that certain cases are counterexamples to my view.

This, I hope, serves as a clear articulation of the way in which I approach apparent borderline cases and apparent counterexamples generally, having provided a theoretical elaboration for what is, after all, a philosophical term of art. To put it in a slogan: *although intuitions are welcome to the party, the theory buys the drinks*. With that, I am ready to play the assertion counterexample game.

5.2. **Objection 1: bald-faced lies**

Consider the following cases:

Loyal Butler

> A gifted pianist is practicing when the telephone rings inconveniently. The butler picks up the phone and says to the caller that the maestro is not home. The caller objects by saying, 'But I hear him playing'. The butler responds, 'You are mistaken, I'm dusting the piano keys'.[138]

[138] Adapted from Sorenson (2007).

Polite Host

> A party guest is wreaking havoc – he is drunk, offending people, bumping into and breaking things, and generally causing a scene. The host suggests to the guest that he should leave. As the guest is leaving, he says, 'I'm sorry, I shouldn't have drunk so much'. The host says politely, 'Not at all. It was an utter pleasure having you. We look forward to seeing you soon'.

Cheating Student

> A Dean catches a student red-handed cheating during an exam. However, there is no direct confession from the student. The Dean has a strict policy by which she refuses to punish any student for cheating in the absence of a confession from that very student – which is well-known by all students. The Dean calls the student into her office and plays video footage of the incident. She then asks, 'Did you cheat?' The student says, 'No, I did not cheat'. The Dean sends the student away unpunished.[139]

Cases of this sort are known as 'bald-faced lies'. These types of cases are described as ones in which the speaker asserts that *p* even though it is common knowledge (between speaker and hearer) that the speaker does not believe that *p*. That is, the speaker and the hearer each know that the speaker does not believe that *p*, and each know that each know that the speaker does not believe that *p*, and each know that each know that each know that the speaker does not believe that *p*, and so on. It is this last point specifically, i.e., it being common knowledge (between speaker and hearer) that the speaker does not believe that *p*, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*. Thus, we seem to have an argument against (AE):

---

[139] Adapted from Carson (2006).

Premise 1: If so-called 'bald-faced lies' are assertions, then it is possible to make an assertion without intending that the hearer acquire a belief that *p.*

Premise 2: If it is possible to make an assertion without intending that the hearer acquire a belief that *p*, then (AE) is false.

Premise 3: So-called 'bald-faced lies' are assertions.

Conclusion: (AE) is false.[140]

By using a novel adaptation of Moore's paradox based on my theoretical elaboration of our candidate explicatum, I intend to show, however, that premise 3 is false. Note, it is traditionally thought that it is a necessary condition on lying that the speaker intends that the hearer acquire a belief in a proposition which the speaker believes to be false. As such, many have argued that so-called 'bald-faced lies' show that the traditional definition of lying is false.[141] Therefore, my defence of (AE) also functions as a defence of the traditional definition of lying.

5.2.1. **Moorean Paradox**

Moore writes:

> I start from this: that it's perfectly absurd or nonsensical to say such things as ['*p* but I don't believe that *p*']. So far as I can see, when we say it's absurd to *say* ['*p* but I don't believe that *p*'], we don't mean that it's absurd merely to utter *the words*, but something like this: it's absurd to say them in the sort of way in which people utter sentences, when they are using these sentences to *assert* the proposition which these sentences express. I will call this 'saying them assertively'. I don't want to say that to

[140] Those who defend the view that so-called 'bald-faced lies' are genuine assertions and thus genuine lies include Carson (2006), Fallis (2009; 2012; 2014), Garcia-Carpintero (2018), Saul (2012), Sorensen (2007), and Stokke (2013; 2018).

[141] See Mahon (2014).

> utter sentences assertively is the same as making an assertion. (Moore in Baldwin (ed.) 2003: 207)

He writes further:

> [W]hat I imply by saying [‘*p* but I don’t believe that *p*’] contradicts what is meant by my saying that I don’t believe it: so that there is a contradiction between what I implied and something I said, though none between the two things I said. And I suggest that this is why it is absurd for me to say it, even though there’s no contradiction between the two things I said. But now the question may be raised: What’s meant by saying that I imply it? The only answer I can see to this is that it is something which follows from the following empirical fact: *viz*. that in the immense majority of cases in which a person says a thing assertively, he does *believe* the proposition which his words express. (Moore in Baldwin (ed.) 2003: 210)

It is worth making a few comments here. First, the reason why such cases are considered paradoxical is because there is nothing problematic in both *p* being true and an individual not believing that *p*, but seemingly there is something problematic with a speaker asserting both at the same time. Second, Moore claims that when a speaker makes an assertion that *p*, they ‘imply’ that they believe that *p*. His account of the way in which assertion implies belief rests on an inductive association, whereby the link between assertion and belief is so common that we experience utterances in which this implication is contradicted as wrong.[142] However, we need not endorse this Moorean notion of ‘implication’. Why? Because we can account for the sense in which a speaker, when they make an assertion, not only expresses or manifests a belief in what they assert, but indeed represents themselves as having the knowledge in question, by appealing to the maxims of quality, i.e., the specific mutual default expectation that speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that

[142] See Moore in Baldwin (2003).

what the speaker asserts, *p*, is true is a corollary. Third, and especially important for our purposes, Moore makes a point of distinguishing between 'uttering a sentence assertively' and 'making an assertion'. I submit that the most plausible reading of Moore's distinction is this: although a speaker may utter assertively '*p* but I don't believe that *p*', they will not succeed in making an assertion to that effect in doing so. Note, though, my defence of (AE) does not depend on whether Moore actually endorses this particular reading of his distinction or not, nor does it depend on any pre-theoretical intuitions I may have about his distinction. Rather, the fact that I am starting from my theoretical elaboration of our candidate explicatum vindicates this particular reading of Moore's distinction as the most plausible, as I will show.[143]

5.2.2. **A Gricean Account of Moorean Paradox**

Consider the following:

Bank Location Frustration

> Shannon is new in town and does not know where the bank is, so she asks a passing stranger, Sylvester:
>
> Shannon: Excuse me, is the bank on the East- or West-side of town?
> Sylvester: It's on the East-side of town, but I don't believe that it's on the East-side of town.
> Shannon: …?

[143] As such, any objections against this particular reading of Moore's distinction which do not respect the fact that I am using my theoretical elaboration of our candidate explicatum as our starting point may be seen as dialectically illegitimate.

Shannon is at a complete loss. She is not able to take Sylvester's utterance as a joke, she is not able to calculate an implicature from it, and she is not able to ascribe the belief that the station is on the East-side to Sylvester, so that she can interpret and, in turn, comprehend him as having asserted anything to her. In sum, Sylvester's utterance does not play a proper role in the conversation. What we have here is a standard Moorean paradox.

Using the theoretical elaboration of our candidate explicatum, we can diagnose this case in the following way. Sylvester utters assertively, which is to say that he makes his utterance in such a way which would usually license Shannon to ascribe to Sylvester the intention that she acquire a belief on the basis of her recognising his intention that she acquire said belief. However, Sylvester utters assertively not only that *p*, but also his lack of belief in *p*, openly flouting the maxims of quality, i.e., the specific mutual default expectation that speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary. Since Shannon is not able to employ the maxims of quality, she is not able to interpret and, in turn, comprehend Sylvester's behaviour as an act of assertion. This thus results in Sylvester apparently making a contribution (in virtue of performing a locution), but not successfully making a contribution (failing to perform an illocution) such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the exchange in which he and Shannon are purportedly engaged. In other words, Sylvester blocks the ways in which Shannon tries to recognise his intention, rendering his utterance wholly uncooperative.[144]

Austin's (1962) discussion of 'infelicities' is instructive here, in which he talks about the 'unhappy' functioning of a speech act. According to Austin, infelicities fall into two categories, since there are two ways in which the functioning of a speech act may be unhappy:

---

[144] Note, the case could be reformulated such that Sylvester utters assertively that he believes that not-*p* rather than not believing that *p*, accounting for both the 'omissive' and 'commissive' versions of Moorean paradox.

> Abuses: the speech act is achieved but it is 'insincere'.
> Misfires: the speech act is not achieved, because the speech act which the speaker purports to perform is 'botched'. (13–24)

As I see it, Moorean paradox is best understood in terms of Austinian misfire, not abuse.[145] Indeed, understanding Moorean paradox in terms of Austinian misfire provides a neat way of understanding Moore's distinction between 'uttering assertively' and 'making an assertion'. Accordingly, let us call Moorean paradox an instance of 'botched assertion', where a botched assertion should *not* be seen as an instance of assertion, but rather as an instance of a special type of utterance characterised by its failure to be the type of speech act it purports to be.[146]

5.2.3. **Contextual Moorean Paradox**

What can we learn from Moore's paradox? I suggest the following:

> Moorean lesson: If a speaker utters assertively that *p* whilst at the same time (at least apparently) *openly* not believing that *p*, the speaker's utterance will result in Moorean absurdity.

---

[145] Thank you to Mitch Green (personal correspondence) for a helpful conversation about this.

[146] Another (albeit non-Moorean) case in which speakers' assertions might be thought to misfire is in a case put forward by Kripke (1975), where the professor in classroom 1A writes on the board, 'Whatever the professor in 1B writes on their board is true', and where the professor in classroom 1B at the same time just happens to write on their board, 'Whatever the professor in 1A writes on their board is false'. As I see it, both professors' assertions misfire, purely by coincidence.

By fleshing out the notion of openness in terms of 'mutual manifestness', we can gain a better understanding of what Moorean absurdity amounts to.[147] Specifically, Moorean absurdity is the clash between the mutual manifestness of the speaker's lack of belief that *p* and the specific mutual default expectation that the speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary. Since the employment of said mutual default expectation is essential to comprehending the speaker's intention that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*, the clash between the mutual manifestness of the speaker's lack of belief that *p* and said mutual default expectation explains why the speaker's intention is thwarted. Moreover, this provides us with a clear articulation of the way in which Moorean paradox results in botched assertion. What is especially interesting about this, though, is that a speaker's (at least apparent) open lack of belief can manifest itself in more ways than Moore appreciates. Specifically, it can manifest itself in a speaker and hearer's common knowledge. Consider the following:

Puzzling Poltergeist

> It is common knowledge between friends Jack and Jane that Jack believes in ghosts and that Jane does not – he is head of the *Paranormal Activity Society* and she is head of the *Shifty Sciences Society*. Jack's wife calls when he is with Jane and, after the call, he says, 'My wife's angry because she found the front door unlocked when she came home. Now I think about it, it was probably my son, whom I suspect went out after me but before my wife got in'. Jane says, 'It can't have been him. He was at our house'. Jack realises Jane is correct. Jane takes a sip of her drink and says, 'So, it must

[147] See Sperber & Wilson (1986) for more on mutual manifestness. Essentially, something is manifest to an individual if and only if they can represent it mentally and accept its representation as true or probably true. Mutual manifestness therefore is when something is manifest, as just described, between a speaker and a hearer, mutually.

have been ghosts'. Jack scowls playfully at her, 'Alright, quit mocking me'. 'I'm not mocking you. It was ghosts', she says seriously. Jack responds confused and irritated, 'Please, stop'. Jane looks at Jack earnestly and says, 'It's ghosts that unlocked your front door, trust me'. Jack stares at Jane, puzzled.

Jack is at a complete loss. He is not able to take Jane's utterance as a joke, he is not able to calculate an implicature from it, and he is not able to attribute the belief that ghosts unlocked the door to her, so that he can interpret and, in turn, comprehend her as having asserted anything to him. In sum, Jane's utterance does not play a proper role in the conversation. What we have here is a *contextual* Moorean paradox. [148]

Again, using the theoretical elaboration of our candidate explicatum, we can diagnose this case in the following way. Jane utters assertively, which is to say that she makes an utterance in such a way which would usually license Jack to ascribe to Jane the intention that he acquire a belief on the basis of him recognising her intention that he acquire said belief. However, Jane utters assertively that *p* when it is common knowledge (between Jack and Jane) that Jane does not believe that *p*, openly flouting the maxims of quality, i.e., the specific mutual default expectation that speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary. Since Jack is not able to employ the maxims of quality, he is not able to interpret and, in turn, comprehend Jane's behaviour as an act of assertion. This thus results in Jane apparently making a contribution (in virtue of performing a locution), but not successfully making a contribution (failing to perform an illocution) such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the exchange in which she and Jack are purportedly engaged. In

[148] Max Black (1952) serves as my inspiration here. He gives an example of a speaker uttering assertively that they are 2ft tall when it is common knowledge (between speaker and hearer) that the speaker is at least 6ft tall, and their assertion being 'inoperative' as a result.

other words, Jane blocks the ways in which Jack tries to recognise his intention, rendering her utterance wholly uncooperative, thus resulting in a botched assertion.

5.2.4. **Contextual Moorean Paradox Test**

Apart from being interesting in and of itself, the contextual Moorean paradox can be put to work. Consider the following:

Contextual Moorean Paradox Test

> If a speaker utters that *p* when it is common knowledge (between the speaker and the hearer) that the speaker does not believe that *p*, there are only two options regarding the status of the speaker's utterance:
>
> a) The speaker performs some (speech) act other than assertion, or
> b) The speaker's utterance results in Moorean absurdity (i.e., botched assertion).

How does the test work? To start, consider any case of a speaker uttering that *p* and re-construct it (if it is not already so constructed) in such a way that the speaker's utterance is performed against a context in which it is common knowledge (between the speaker and the hearer) that the speaker does not believe that *p.* Then, make a verdict on whether the utterance results in Moorean absurdity or not. This will depend on the plausibility of ascribing to the speaker the intention that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*, which itself will depend on the plausibility that the specific mutual default expectation that speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary is in place. If the utterance does not result in Moorean absurdity, then it must be some (speech) act other than assertion. If the utterance does result in Moorean absurdity, then it must be a botched assertion. The crucial observation to make here, though, is that whether we go with option (a) or option (b), the speaker has not made an assertion. Thus,

the contextual Moorean paradox test demonstrates that so-called 'bald-faced lies' are not assertions. As such, premise 3 is false. (AE) is preserved.

5.2.5. **The Many Faces of So-called 'Bald-faced Lies'**

Following Keiser (2016), I not only hold the view that so-called 'bald-faced lies' are not assertions, but I also hold the view that the term 'bald-faced lie', as it has been used in the philosophy literature, fails to pick out a single act type. Indeed, I think that cases of so-called 'bald-faced lies' cut across options (a) and (b) from the contextual Moorean paradox test. Let us consider the ways in which we might interpret cases of so-called 'bald-faced lies' which seem to fall most comfortably under category (a): the speaker performs some (speech) act other than assertion. Perhaps some cases of so-called 'bald-faced lies' are best understood as cases of implicature, whereby the speaker utters that *p* in a context in which it is common knowledge (between the speaker and the hearer) that the speaker does not believe that *p* in order to implicate *q*. (Notice that *Loyal Butler* may be plausibly interpreted in this way.) Perhaps some cases of so-called 'bald-faced lies' are best understood as cases of 'acting', whereby the speaker utters that *p* in a context in which it is common knowledge (between the speaker and the hearer) that the speaker does not believe that *p*, because the speaker and the hearer are engaged in some form of mutual pretence.[149] (Notice that *Polite Host* may be plausibly interpreted in this way.) Or, perhaps some cases of so-called 'bald-faced lies' are best understood as cases of '(quasi-)institutional' speech acts, whereby the speaker utters that *p* in a context in which it is common knowledge (between speaker and hearer) that the speaker does not believe that *p* in order to, say, formally 'enter a plea' of not guilty. (Notice that *Cheating Student* may be plausibly interpreted in this way.)[150]

Let us move on now to cases of so-called 'bald-faced lies' which seem to fall most comfortably under category (b): the speaker's utterance results in Moorean absurdity (i.e.,

---

[149] Maitra (2018) makes an argument for a position which is like this.

[150] Keiser (2016) makes an argument for a position which is like this.

botched assertion). Cases of so-called 'bald-faced lies' which fall under category (b) arise for various reasons. For example, perhaps the speaker is conversationally cornered and wishes to opt out of the exchange, and yet cannot do anything but perform a botched assertion given the other conversational moves the exchange allows them to make.[151] A particularly interesting type of case which falls under category (b), however, is when a speaker deliberately and nefariously performs a botched assertion. Why would a speaker deliberately and nefariously perform a botched assertion? Because the speaker wishes to put the hearer in the position where they cannot judge properly what the speaker is doing, in an attempt to tease, afront, intimidate, or confuse the hearer for their own gain or amusement. In other words, these are cases in which the speaker is 'trolling' the hearer.[152] As I see it, such cases are best understood in terms of the notion of 'conversational perversion', a notion recently developed by Bourne & Caddick Bourne (2018).[153]

5.2.6. **So-called 'Bald-faced Lies' & Conversational Perversions**

Bourne & Caddick Bourne's inspiration for their account of conversational perversion comes from Nagel's (1969) account of sexual perversion. Nagel suggests understanding sexual perversion as something which thwarts the reciprocal awareness and recognition involved in what he calls a 'complete sexual encounter'. He writes:

---

[151] Politicians and public figures are often seen behaving as such, especially when conversationally cornered by journalists.

[152] Just to be clear, I am not trying to define what trolling is here, nor am I committed to the position that trolling must be understood solely in terms of Bourne and Caddick Bourne's notion of conversational perversion. Rather, I am merely illustrating how Bourne & Caddick Bourne's notion of conversational perversion helps articulate the mechanisms at play in at least some instances of trolling. For more on trolling, see Barney (2016).

[153] Meibauer (2014) argues that so-called 'bald-faced lies' are acts of 'verbal aggression'. By understanding so-called 'bald-faced lies' in terms of conversational perversion, we have a neat way of articulating the underlying mechanisms at play in such acts of verbal aggression.

> [R]eflexive mutual recognition is to be found in the phenomenon of meaning, which appears to involve the intention to produce a belief or other effect in another by bringing about his recognition of one's attention to produce that effect. (That result is due to H. P. Grice, whose position I shall not attempt to reproduce in detail.) Sex has a related structure: it involves a desire that one's partner be aroused by the recognition of one's desire that he or she be aroused. (Nagel 1969: 12)

Although Nagel's observation is made in passing, Bourne & Caddick Bourne claim that it is key to articulating the possibility of perversion not only in the sexual but in the conversational case. The structure Nagel identifies is a sexual analogue of the telling mechanism. Call it the 'Nagelian mechanism':

> Nagelian mechanism: *A* desires that *B* become aroused on the basis of *B* recognising *A*'s desire that *B* become aroused.[154]

The idea here is that a sexual act is perverted if it does not allow for the instantiation of the Nagelian mechanism. For instance, an act that involves selecting a sexual partner which one thinks is incapable of recognising one's desire is perverted for that very reason (e.g., bestiality). Bourne & Caddick Bourne's idea, then, is to use the structural similarity between conversational and sexual encounters to define various conversational perversions corresponding to sexual perversions. One type of perversion that they focus on is sadism. According to them, the way to pinpoint the source of the perversion in sexual sadism is to focus on what is needed in order for a person to recognise a desire that they be aroused: the desire must be *demonstrated* to them. They write:

> Sadistic actions do not *demonstrate* a desire that the other person be aroused. Instead, they demonstrate a desire that the partner suffer or that they be hurt or humiliated (for example), rather than that they be aroused. By not *demonstrating* the desire that

[154] See Bourne & Caddick Bourne (2018: 148).

> the other person be aroused, the sadist blocks the other person from recognising that desire in them. Thus in choosing sadistic acts, a person does not desire that they demonstrate, to their partner, the desire that the partner be aroused, and so does not desire that the partner *recognise* a desire that they be aroused. The desire that the partner be aroused on the basis of recognising the desire that they be aroused is therefore missing, and the basic Nagelian mechanism is not instantiated. (2018: 149—150)

With this in mind, recall that the telling mechanism captures a structure in which conversational participants comprehend each other by means of mutual recognition of the speaker's intention, whereby the speaker intends that the hearer acquire a belief that *p* on the basis of the hearer recognising the speaker's intention that the hearer acquire a belief that *p*. Moreover, recall that the ways in which conversational participants attribute such reflexive intentions to each other is described by appealing to our communicative standards, i.e., the cooperative principle, and thus the conversational maxims which fall within it. Bourne & Caddick Bourne's idea, then, is that our communicative standards can be exploited by a speaker to *deliberately* block the ways in which the hearer tries to recognise the speaker's reflexive intention, in a way that parallels sexual sadism. Consider the following:

Jargon Juggernaut

> Olivia, a project manager for a large organisation, has put together an initial plan for the deployment of a new IT system. However, she has not consulted the IT team. During a presentation of her proposed strategy, one of the IT experts in the audience, Drew, realises that several steps will not work, because Olivia has failed to consider certain technical limitations. Irritated by her lack of communication with the IT team, Drew raises his hand and, when invited to speak, reels of the following: 'If you wish to deploy this new IT system, you need to decide whether it is going to be an on-premise or cloud-hosted solution, and then, depending on that, you need to work out how it is going to fit with our virtualised SQL infrastructure, especially given the complications of our DMZ setup, while at the same time complying with our N3

> code of connection'. Olivia stands in silence before the audience, confused and embarrassed.

The idea in this particular case is that Drew deliberately utters things which he knows Olivia does not have the expertise to understand, by choosing obscure specialist terminology. What Drew chooses to utter is selected precisely because Olivia cannot identify what beliefs she should acquire and thus cannot respond suitably in the conversation. Bourne & Caddick Bourne call this an instance of 'conversational sadism':

> The sadist does not intend that the hearer acquires a belief on the basis of recognising their [reflexive] intention, but rather intends that the hearer feels they cannot judge what the sadist's [reflexive] intentions are. (Bourne & Caddick Bourne 2018: 150)

Now, I should note that Bourne & Caddick Bourne do not apply the notion of conversational perversion to cases of so-called 'bald-faced lies'. However, so-called 'bald-faced lies' seem to fit the model well. Specifically, they seem to be a unique type of conversational sadism. That is, rather than, say, deliberately using obfuscating technical jargon to block the instantiation of the telling mechanism, so-called 'bald-faced liars' block the instantiation of the telling mechanism by deliberately uttering assertively what they openly disbelieve. To illustrate this, consider again *Puzzling Poltergeist* which seems to be a case of so-called 'bald-faced lying' which falls comfortably under category (b) from the contextual Moorean paradox test. Jane utters assertively, which is to say that she makes an utterance in such a way which would usually license Jack to ascribe to Jane the intention that he acquire a belief on the basis of him recognising her intention that he acquire said belief. However, Jane *deliberately* utters assertively that *p* when it is common knowledge (between Jack and Jane) that Jane does not believe that *p*, *deliberately* openly flouting the maxims of quality, i.e., the specific mutual default expectation that speaker asserts that *p* in virtue of the fact that the reason why they assert that *p* is that they know that *p*, of which the super mutual default expectation that what the speaker asserts, *p*, is true is a corollary. Since Jack is not able to employ the maxims of quality, he is not able to interpret and, in turn, comprehend Jane's behaviour as an act of assertion. This thus results in Jane apparently making a contribution

(in virtue of performing a locution), but *deliberately* not successfully making a contribution (*deliberately* failing to perform an illocution) such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the exchange in which she and Jack are purportedly engaged. In other words, Jane *deliberately* blocks the ways in which Jack tries to recognise his reflexive intention, rendering her utterance wholly uncooperative, thus *deliberately* resulting in a botched assertion.

5.2.7. **Challenging So-called 'Bald-faced Liars'**

Although Bourne & Caddick Bourne do not explore the idea of challenging conversational perverts, it seems an interesting topic. So-called 'bald-faced liars' are particularly interesting in this respect, because they are sometimes mistakenly thought to be performing genuine assertions and thus genuine lies. A such, a hearer faced with a so-called 'bald-faced lie' might readily, though uneasily, respond to the speaker the only way they know how. For example, they might say, 'That is false!', 'You don't believe that!', or 'How do you know that?!'. Of course, such epistemic challenges are misguided, since it is common knowledge (between the speaker and the hearer) that the speaker disbelieves what they utter – that is the crux of so-called 'bald-faced lies', after all. However, given the enigmatic nature of so-called 'bald-faced lies', such responses are understandable and perhaps not uncommon. Unfortunately, if a hearer does take this line against a so-called 'bald-faced liar', they will have been led astray, which may create frustration, confusion, and embarrassment in the hearer, resulting in the speaker's amusement or gain. Thus, we can further see the sense in which so-called 'bald-faced lies' are an appropriate tool for trolling. How, then, should we go about challenging so-called 'bald-faced liars'? The obvious answer is that we should charge them with being wholly uncooperative conversationally. However, before we rush to declare victory over so-called 'bald-faced liars', we must remember that such conversational perverts are decidedly brazen in their behaviour – they are overt with respect to the infelicity of their utterances. For instance, political journalist Andrew Sullivan (2019) recalls a television interview between Donald Trump and George Stephanopoulos in which Trump says, 'I like the truth. I'm actually a very honest guy'. According to Sullivan, Trump's utterance 'was so staggeringly, self-evidently untrue, and so confidently, breezily said, it

was less a statement of nonfact than an expression of pure power'.[155] As such, even if we challenge so-called 'bald-faced liars' as being wholly uncooperative conversationally, such a challenge might be just as futile as the epistemic challenges mentioned above. Indeed, the speaker may simply respond with more so-called 'bald-faced lies', further demonstrating their brazenness. It is no coincidence that advice for ignoring trolls generally is, 'Please don't feed the trolls', after all.[156] Nonetheless, if we want to hold people accountable, we need to know what it is we are holding them accountable for.[157, 158] So, we can talk in terms of 'bald-faced lies' if we like, but not only should we not buy into the idea that they are genuine assertions and thus genuine lies for purely philosophical purposes, we should not buy into this idea for social and political purposes as well.

5.3. **Objection 2: double bluffs**

Consider the following case:

Sneaky Spy

> Suppose a spy, Natasha, asserts to her foreign counterpart that her partner is dead, yet she knows full well that he knows that her partner is alive and is being held by

---

[155] See Sullivan (2019).

[156] See Sullivan (2012).

[157] At this point, it might be thought that *Cheating Student* falls more comfortably under category (b) from the contextual Moorean paradox test (in §5.2.4), rather than category (a). I certainly think that this interpretation of the case has legs, for when the student is called into the Dean's office and confronted by the Dean about having cheated, the student could avoid punishment by simply staying silent or responding with, 'No comment' – only a confession will lead to their punishment, after all. However, the student decides not only to avoid punishment but to troll the Dean in the process by uttering assertively something it is common knowledge (between the student and the Dean) that the student believes to be false, namely, that they did not cheat.

[158] Although it is not a case of so-called 'bald-faced lying' but rather a case of standard Moorean paradox, notice that we can also interpret Sylvester's behaviour in *Bank Location Frustration* as a case of conversational perversion and thus as a case of trolling.

him in a nearby facility. Suppose, though, that the reason she asserts to him that her partner is dead is because she intends him to believe that she believes her partner is dead.[159]

Let us call cases of this sort 'double bluffs'. This type of case is described as one in which the speaker asserts that *p*, the speaker believes that not-*p*, and the speaker knows that the hearer knows that not-*p*. (Notice that, unlike cases of so-called 'bare-faced lies', it is not common knowledge (between speaker and hearer) that the speaker believes that not-*p*, since it is not the case that the speaker and the hearer know that the speaker knows that not-*p*.) It is this last point, i.e., that the speaker knows that the hearer knows that not-*p*, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*.[160] Thus, we seem to have another argument against (AE):

Premise 1:	If so-called 'double bluffs' are assertions, then it is possible to make an assertion without intending that the hearer acquire a belief that *p*.
Premise 2:	If it is possible to make an assertion without intending that the hearer acquire a belief that *p*, then (AE) is false.
Premise 3:	So-called 'double bluffs' are assertions.
Conclusion:	(AE) is false.

By offering up an alternative interpretation of *Sneaky Spy* based on my theoretical elaboration of our candidate explicatum, I intend to show, however, that premise 3 is false. To appreciate the thinking behind my alternative interpretation of *Sneaky Spy*, it will be helpful to keep in mind what Natasha's overarching aim is in this case, namely, for her

---

[159] Case inspired by Searle (1969).

[160] It might be thought that it is possible for a speaker to make an assertion that *p* to a hearer who believes or knows that not-*p* and thereby intend that the hearer acquire a belief that *p*, so that the hearer relinquishes their belief and/or loses their knowledge that not-*p*. I certainly think this is possible. However, the stipulation in *Sneaky Spy* is that it is not such a case. My argument against *Sneaky Spy* respects this stipulation.

foreign counterpart to acquire a belief that Natasha herself believes that her partner is dead. Here, then, is my alternative interpretation of the case.

Natasha utters assertively, which is to say that she makes an utterance in such a way which would usually license her foreign counterpart to ascribe to Natasha the intention that they acquire a belief that her partner is dead on the basis of them recognising Natasha's intention that they acquire a belief that her partner is dead. What is more, Natasha's foreign counterpart does indeed ascribe this intention to Natasha. However, they do so mistakenly, for Natasha does not actually have this intention. Thus, she does not actually make an assertion that her partner is dead. Instead, the point of Natasha's utterance is to give her foreign counterpart the *impression* that she has made an assertion that her partner is dead and thus giver foreign counterpart the *impression* that she intends that they acquire a belief that her partner is dead on the basis of them recognising Natasha's intention that they acquire a belief that her partner is dead. Why? Because in doing so she gives her foreign counterpart the *impression* that the specific mutual default expectation that Natasha asserts that her partner is dead in virtue of the fact that the reason why she asserts that her partner is dead is that she knows that her partner is dead, of which the super mutual default expectation that what she asserts, that her partner is dead, is true is a corollary, is instantiated between them when it is not. (Of course, this is conditional on Natasha giving the *impression* that the mutual default expectation that she is making a contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the exchange in which she and her foreign counterpart are *purportedly* engaged is instantiated between them when it is not.) Since Natasha's foreign counterpart, however, knows the true status of her partner, Natasha thereby gives her foreign counterpart the *impression* that she satisfies only a certain corollary of said *purported* mutual default expectations, namely, that she believes (albeit falsely) that her partner is dead when she does not. In other words, Natasha performs a fake assertion, not a real assertion. (AE) is preserved.

What is attractive about this interpretation of *Sneaky Spy* is not only that it preserves (AE), but that it also provides a clear account of the mechanism of deception at play in double bluffs. Deception is commonly understood to involve one individual intentionally causing

another individual to acquire a belief that *p* when they themselves believe that *p* is false.[161] This of course is precisely what Natasha is doing, but she is not doing it by means of communication (e.g., the telling mechanism). Instead, she is doing it by giving her foreign counterpart the *impression* that she is so communicating, with the overarching non-communicative intention that her foreign counterpart acquire a belief that Natasha believes that her partner is dead. Strictly speaking, then, Natasha is not lying to her foreign counterpart, though she nonetheless sets out to deceive them.[162]

5.4. **Objection 3: resolute assertions**

Consider the following case:

Interrogation Room

> In the course of their interrogation by the police, it must have become clear very quickly to the members of the Birmingham Six that nothing they could say or do would persuade their interlocutors either that they (the suspects) had not planted the bombs or that they (the suspects again) believed that they had not planted the bombs. For all that, when they uttered the words 'We did not plant the bombs', the suspects certainly asserted as much.[163]

---

[161] See Chisholm & Feehan (1977).

[162] To draw out the difference between a case of lying and a case of non-communicative deception, compare a speaker who asserts to a hearer that they are sick when they themselves believe this to be false against a speaker who makes an assertion about anything other than being sick but does so in such a way (i.e., with a certain look, voice, cough, splutter, etc.) which indicates to those around them that they are sick when they themselves believe this to be false. Both are instances of attempted deception. Strictly speaking, though, the former is a lie whereas the latter is not.

[163] Adapted from Rumfitt (1995).

Let us call cases of this sort 'resolute assertions'. This type of case is described as one in which the speaker asserts that *p* whilst at the same time knowing that the hearer will not believe what they assert. It is this last point, i.e., that the speaker knows that the hearer will not believe what they assert, which supposedly determines why the speaker does not – because they cannot – intend that the hearer acquire a belief that *p*. Thus, we seem to have another argument against (AE):

Premise 1:	If so-called 'resolute assertions' are assertions, then it is possible to make an assertion without intending that the hearer acquire a belief that *p.*
Premise 2:	If it is possible to make an assertion without intending that the hearer acquire a belief that *p*, then (AE) is false.
Premise 3:	So-called 'resolute assertions' are assertions.
Conclusion:	(AE) is false.

By offering up separate alternative interpretations of *Interrogation Room* based on my theoretical elaboration of our candidate explicatum, I intend to show, however, that either premise 1 or premise 3 is false, depending on which interpretation we ultimately opt for. To appreciate the thinking behind my separate alternative interpretations of *Interrogation Room*, it will be helpful to keep in mind the underlying intuition that drives the objection, namely, that an individual cannot intend to do something which they themselves know to be impossible to do. Here, then, is my first alternative interpretation of the case. I start with a general question from Thalberg (1962):

- When an individual recognises the overwhelming impediments to an enterprise, can they nonetheless intend to accomplish it?[164]

[164] Rephrased slightly for clarity. See Thalberg (1962: 49).

As he notes, the answer to this general question depends on the answers to the following two specific questions:

- Is the notion of 'knowing it is impossible' always clear enough to serve as a limiting case of intending?
- If an individual can try to reach goals which they know to be unattainable, should they ever be understood as intending to succeed in such attempts?[165]

Thalberg answers the first specific question negatively and the second specific question affirmatively, in light of the following example:

> Imagine a heroin addict who yearns to free himself of his habit: he knows, from having failed in earlier attempts to resist his craving, and from having studied the case-histories of other addicts, that he has practically no chance of accomplishing the feat. Suppose that the addict makes a sedulous attempt to cease taking his anodyne, and […] manages to quell his agonizing desire for the drug. If his friends congratulate him on his amazing reformation, should he declare, 'Don't praise me; I tried, but I didn't really intend to stop taking narcotics'? His statement expresses more than modesty; it borders on paradox. (55–56)

Thalberg maintains therefore that there are at least some cases in which an individual recognises the overwhelming impediments to an enterprise but nonetheless intends to accomplish it.[166] Furthermore, he is not alone in thinking this, for his view is shared by Anscombe (1957). She writes:

---

[165] Rephrased slightly for clarity. See Thalberg (1962: 49).

[166] Of course, as Thalberg gestures, most people strive for realisable goals. Moreover, we tend to rely more often on informed individuals to accomplish what they intend. However, individuals who are, say, uninformed, anxious, or even in some sense irrational nonetheless have intentions.

> 'I am going to […] unless I do not' is not like 'This is the case, unless it isn't […]' In some cases one can be as certain as possible that one will do something, and yet intend not to do it […] A man could be as certain as possible that he will break down under torture, and yet determined not to break down. (93)

Given these considerations, I submit that the suspects in *Interrogation Room* may plausibly be interpreted as resolutely making an assertion that they did not plant the bombs and thereby trying to get their interlocutors to believe that they did not plant the bombs (and thus intending as much), even though they apparently 'know it is impossible'. If so, premise 1 is false. (AE) is preserved. Let us turn now to my second alternative interpretation of the case.

My second alternative interpretation of *Interrogation Room* is a development on the plausible idea (mentioned in §3.7) that institutions can perform speech acts. However, I am not interested here in the speech acts that institutions perform, but rather I am interested in the speech acts that institutions themselves can comprehend. After all, if it is plausible that institutions can perform their own speech acts, it seems unreasonable not to accept that they can comprehend speech acts from others as well. Here is what I have in mind. The way that *Interrogation Room* is often presented suggests that the suspects are engaging solely with the interrogators who are in the room with them. But it seems plausible that they are instead engaging with the larger institution of which the interrogators are part. In the case of the Birmingham Six, this would be West Midlands Police, if not the whole of UK law enforcement. That is, strictly speaking, the larger institution itself is the suspects' interlocuter, not the interrogators as such. The idea is that in just the same way that some single member of an institution may act as a 'mouthpiece' for said institution's assertions (regardless of the fact that said member personally does not have the knowledge in question), a single member of an institution may seemingly act as an 'earpiece' for the assertions the institution receives (regardless of the fact that the speaker does not – because they cannot – intend said member personally to acquire the relevant belief). Given these considerations, I submit that the suspects in *Interrogation Room* may plausibly be interpreted as resolutely making an assertion that they did not plant the bombs and thereby intending

that the institution (not the interrogators in the room with them necessarily) believe that they did not plant the bombs. If so, then premise 1 is false. (AE) is preserved.

The two alternative interpretations of *Interrogation Room* which I have provided so far respect the stipulation that the suspects make an assertion that *p*, though both alternative interpretations of course aim to show that the speakers should also be understood as thereby intending that the hearer (when it is properly understood who the hearer is) acquire a belief that *p*. Thus, they are objections to premise 1. My third and final alternative interpretation of *Interrogation Room*, however, takes a different tack. By utilising Keiser's (2016) work on 'language games', I intend to show that when the suspects utter to the interrogators, 'We did not plant the bombs', they may plausibly be interpreted as performing some act other than assertion.[167] If so, premise 3 is false. (AE) is preserved.

According to Keiser, there are many varied activities in which we employ language, yet not all of them qualify as conversations. Those activities in which we employ language, but which do not qualify as conversations, she categorises as language games.[168] As such, she thinks it is possible for a speaker to make a move in a language game without making a move in a conversation. Crucially, though, she thinks that the speech act of assertion can only occur as part of a conversation, not a language game. To illustrate what a language game is, she provides two key examples. The first is the 'exam game'. Suppose a student is sitting an exam as part of their secondary school education. Keiser's thought is that when the student submits their answers in the exam, they are not engaged in a conversation, but rather they are playing a language game, namely, a language game the goal of which is to

---

[167] Note, Keiser's work on language games is originally discussed in relation to so-called 'bald-faced lies', namely, with the aim of showing that so-called 'bald-faced lies' are not genuine instances of lying because they are not genuine instances of assertion. Indeed, I flag her work in relation to so-called 'bald-faced lies' in fn.150 when discussing possible interpretations of *Cheating Student*. However, it seems to me that Keiser's work may also fruitfully be applied to *Interrogation Room*.

[168] Note, Keiser's use of the expression 'language game' corresponds closely with Wittgenstein's (1953).

submit all and only correct answers.[169] The second example she provides is the 'courtroom game'. Suppose a witness takes the stand as part of a trial. Similarly, Keiser's thought is that when the witness presents their case to the courtroom, they are not engaged in a conversation, but rather they are playing a language game, namely, a language game the goal of which is to formally enter evidence which may then be used by the prosecutor and the defence to score points against each other in the trial.[170] Interestingly, Keiser suggests that the courtroom game – and perhaps any kind of language game – is derivative of genuine conversation. This, she claims, would explain why the moves made in a courtroom game appear so much like the types of acts performed in conversations, e.g., assertions. She writes:

> [T]he courtroom game presumably evolved slowly from what were at one point genuine conversations. That is, initially people probably attempted to resolve conflict and establish justice by coming together to have a genuine conversation about who was at fault and what to do about it. But as this procedure became more and more structured, it evolved into a game that – while appearing an awful lot like conversation – actually departs from it in important respects, having its own independent goals and rules. (471–472)[171]

---

[169] Notice that Keiser's construal of examination situations shores up neatly with the comments I make in §3.7 regarding teaching practices, namely, that students are presented with content from a syllabus which has been deemed appropriate/correct by the institution (i.e., the school and its governing bodies).

[170] Katherine Hawley (personal correspondence) has suggested to me that entering evidence on a witness stand is perhaps best understood as having a performative aspect (e.g., like saying 'I do' at a wedding). Her idea is that a witness's utterance has extra significance due to being uttered in that setting, i.e., an institutional setting in which the witness is under oath. This, of course, would explain why there may be legal repercussions for a witness if they do not play the game in line with the rules set down by the courtroom.

[171] Note, Keiser maintains that another reason why moves in the courtroom game look so much like conversational moves is that it is possible for there to be both a genuine conversation and a courtroom game going on simultaneously, and for these to come together and peel apart at various moments throughout the duration of a trial.

Admittedly, Keiser has her own conception of what an assertion is, and she also has her own conception of what a conversation is. However, it is not important that we delve into either of these, for her general picture is sufficiently clear without doing so, and because we can flesh out her general picture further using the theoretical elaboration of our candidate explicatum I have provided, anyway. Recall that I think that the speech act of assertion essentially involves an intention that the hearer acquire a belief by means of the telling mechanism. Moreover, recall that I think the speech act of assertion can only occur as part of a maximally effective exchange of information, which essentially is a communicative exchange the purpose or direction of which is to inform each other of the facts by means of the telling mechanism, i.e., a communicative exchange the purpose or direction of which is to transmit knowledge. Given this, my suggestion is that, just like the exam game and the courtroom game, the 'interrogation game' is not a conversation (or at least not a maximally effective exchange of information, i.e., not a communicative exchange the purpose or direction of which is to transmit knowledge), but rather a language game. What type of language game is it? I submit that it is a language game similar to the courtroom game, the goal of which is to formally enter evidence which may be used by the institution (in this case, West Midlands Police, if not the whole of UK law enforcement) for any relevant legal proceedings.

Just as Keiser suggests with the courtroom game, I think the interrogation game is derivative of genuine conversation (or at least derivative of maximally effective exchanges of information, i.e., communicative exchanges the purpose or direction of which is to transmit knowledge). This would explain why the moves made in an interrogation game appear so much like the types of acts performed in conversations (or at least maximally effective exchanges of information, i.e., communicative exchanges the purpose or direction of which is to transmit knowledge), e.g., assertions. Indeed, it is plausible that the interrogation game also evolved slowly from what were at one point genuine conversations (or at least maximally effective exchanges of information, i.e., communicative exchanges the purpose or direction of which is to transmit knowledge). That is, initially individuals probably attempted to inform each other of the facts for the purpose of gathering information so that they could resolve conflict and establish justice by coming together to have a conversation

(or at least a maximally effective exchange of information, i.e., a communicative exchange the purpose or direction of which is to transmit knowledge). But as this procedure became more and more structured, it evolved into a game that – while appearing an awful lot like conversation (or at least maximally effective exchanges of information, i.e., communicative exchanges the purpose or direction of which is to transmit knowledge) – actually departs from it in important respects, having its own independent goals and rules. Given these considerations, I submit that the suspects in *Interrogation Room* may plausibly be interpreted as performing some act other than assertion, namely, formally entering evidence as part of legal proceedings. If so, premise 3 is false. (AE) is preserved.[172]

Having laid out my separate alternative interpretations of *Interrogation Room*, there is one final observation I would like to make. When discussing cases like *Interrogation Room*, and even certain cases of so-called 'bald-faced lies' (e.g., *Cheating Student*), some philosophers are fond of using a certain phrase to describe the speakers, namely, that they are 'going on record'. Those who are fond of using this phrase seem to think that it in some way establishes that the speakers have made an assertion in such cases. However, the phrase is used promiscuously while not being theoretically robust. In fact, I am aware of only one attempt to provide a theoretical elaboration of the notion of going on record, which is from Stokke (2013; 2018). Stokke endorses the view that what makes an assertion the type of speech act that it is, rather than some other type of speech act, is that a speaker says that *p* and thereby proposes that *p* become part of the common ground (see §2.1.2). The problem with this, however, as Stalnaker himself observes, is that a speaker might say that *p* and thereby propose that *p* become part of the common ground merely to make an assumption that *p*, say, for the sake of argument – which everyone agrees is distinct from making an assertion. Stokke is aware of this problem and so attempts to remedy it by introducing the

[172] If anyone still has any residual or lingering discomfort about the immediate dismissal of secret diary entries and soliloquies as cases of assertion, they might find some comfort in the idea that they can be accounted for further (i.e., on top of being accounted for in terms of hearer-testimony) as alternative types of language games, ones which are yet to receive full philosophical treatment.

distinction between 'official common ground' and 'unofficial common ground'. He bases his distinction on the claim that when propositional content is proposed to become part of the unofficial common ground it is only so proposed temporarily, whereas propositional content proposed to become part of the unofficial common ground is not. Thus, when a speaker makes an assumption, and thereby proposes that *p* become part of the common ground, strictly speaking they are proposing that it become part of the unofficial common ground. Yet, when a speaker makes an assertion, and thereby proposes that *p* become part of the common ground, strictly speaking they are proposing that it become part of the official common ground. According to Stokke, it is only in the latter case that the speaker goes on record, i.e., when a speaker makes an assertion that *p* and thereby proposes to add *p* to the official common ground. There are, however, two issues which arise for Stokke's distinction between the official common ground and the unofficial common ground. First, as Keiser observes, it seems plausible that an individual might make an assumption that *p* and thereby propose to add *p* to the common ground, but not propose to do so temporarily. Indeed, the speaker (and the hearer, for that matter) might be aware that the assumption that *p* is likely to be true, after all. Moreover, it seems plausible that a speaker might assert that *p* with the knowledge that *p* will only ever be temporarily added to the common ground, since it is highly likely to be swiftly rejected by the hearer. In short, the attempt to ground the distinction between official and unofficial common ground in terms of temporality is misguided. Second, Stokke's distinction between official and unofficial common ground seems worryingly ad hoc. That is, it is highly specific to safeguarding his common ground account of assertion and does not serve any other general purpose or have any other antecedent motivation. Taking all this together, I think the notion of going on record is best put aside until a proper theoretical elaboration of it has been provided to thus make it useful.[173]

---

[173] As I see it, the notion of going on record connotes a certain amount of formality and officialness, formality and officialness which more readily accompanies language games in institutional settings compared to conversations (or at least maximally effective exchanges of information, i.e., communicative exchange the purpose or direction of which is to transmit knowledges). If so, when a speaker goes on record, say, by formally entering evidence as part of legal proceedings, it may have nothing to do with assertion. Thus, even if the notion of going on record

5.5. **Embracing the Assertion Counterexample Game**

At the beginning of the thesis (§2.5), I said that our candidate explicatum, (AE), may, in the end, reveal itself to be too restrictive and thus implausible. However, I also said that until a theoretical elaboration of our candidate explicatum is provided, and the apparent borderline cases of assertion scrutinised in light of it, we have no immediate reason to think that (AE) is not suitably restrictive and thus implausible. In this chapter, equipped with my theoretical elaboration of our candidate explicatum, and drawing on other independently plausible theoretical considerations from the philosophy literature, I hope to have shown that (AE) is suitably restrictive and thus plausible. Moreover, in the process of defending (AE), I hope to have shown that playing the assertion counterexample game is not the daunting prospect it is sometimes made out to be when we prepare for it properly. Indeed, I hope to have shown that it can be a useful philosophical tool, one which not only helps re-align our intuitions, but one which enables us to gain a better understanding of various philosophically interesting phenomena other than assertion, as well assertion itself. Last, but certainly not least, in the course of this chapter, I addressed the final item, item (11), on the substantive list (i.e., Moore's paradox).

---

can in the end be shown to be useful, it is not obvious that it will relate to assertion at all, or at least not exclusively.

6. Conclusion

In the introduction to this thesis, I outlined why assertion is troublesome, namely, that although you will find that talk of it is ubiquitous in philosophy, you cannot always be sure what precisely you have found talk of, or indeed whether you have found talk of anything concrete at all. Moreover, I explained how I have been dogged by this ever since I was an undergraduate student. As such, and having progressed through this thesis, it should be clear why, when I eventually found myself embroiled in the debate on the nature of assertion, I was dissatisfied. That is, it should be clear why it seemed to me that philosophers were often talking past each other and getting caught up in something which was gradually losing significance in the eyes of spectators. However, having now reached the conclusion of this thesis, I hope it is clear that assertion is something which is very much worthy of philosophical attention. Indeed, by attempting to close the gap between the philosophy of language and the epistemology of testimony on the topic of assertion, I hope to have reaffirmed the concept and, in turn, reinvigorated the debate. If so, then the frustration I have felt working with, on, and sometimes against assertion will not have been in vain. More importantly, though, we will have a new way of talking when we talk about assertion. Here is a brief recapitulation.

In *Assertion & Explication*, I introduced the philosophical term of art 'assertion' as it is understood in the philosophy of language and the epistemology of testimony – its main philosophical areas. This included providing an outline of the ongoing debate about the nature of assertion in the philosophy of language and the important though sometimes unclear status assertion has in the epistemology of testimony. I explained that the way in which philosophical discussion about assertion has materialised, especially the lack of integration between the philosophy of language and the epistemology of testimony on the topic, has led to doubts about the effectiveness of the concept. Drawing on Carnap's (1947; 1950) notion of 'explication', however, I held that philosophy simply has not made up its mind about how the concept of assertion is best understood, and that what is needed is a proposal for how it might fruitfully be applied. Specifically, I proposed that assertion is best

understood simply in terms of the paradigm case, and nothing more – this was our candidate explicatum. I acknowledged the potential worry that restricting assertion in the way that I suggested is implausible, given the apparent non-paradigmatic (or borderline) cases of assertion that exist in the philosophy literature. However, I pleaded that our inquiry into the concept should not be guided by such cases. Instead, I recommended that our inquiry should be guided by, connected to, and framed within broader philosophical theories, especially from the philosophy of language and the epistemology of testimony. Essentially, my idea was that we will be best placed to demonstrate the effectiveness of the concept of assertion only when our candidate explicatum has received a full theoretical elaboration.

In *Putting the Epistemology of Testimony to Work*, I called upon the epistemology of testimony to consider the ways in which the speech act of assertion can put an individual in a position to know the truth of what is asserted. I explained that such considerations are apt to guide us in our inquiry and thus lay the groundwork for constructing a theoretical elaboration of our candidate explicatum. Although I showed that there are various ways in which assertion can put an individual in a position to know the truth of what is asserted, I picked out one particular way in which assertion can put an individual in a position to know the truth of what is asserted which I think illuminates the speech act as I understand it: 'knowledge transmission'. As we saw, to fully understand this independently plausible notion from the epistemology of testimony literature, however, we had to understand the nature of the interpersonal relation at play between speakers and hearers in such cases. And, to understand this, we had to understand the conversational dynamics which underpin communicative exchanges generally. I exhibited the current confusion in the epistemology of testimony literature about how to understand these things, the result of which was that I found myself in a predicament. I intended to construct a theoretical elaboration of our candidate explicatum in light of the notion of knowledge transmission. But I did not have an accurate account of the conversational dynamics which underpin communicative exchanges generally. As such, I did not have an accurate account of the nature of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, meant that I

did not have a fully developed view of knowledge transmission in light of which I could construct a theoretical elaboration of our candidate explicatum.

In *A Gricean Account of Assertion*, having first set out my general epistemological commitments, I used, revised, and developed Grice's (1957; 1969; 1975; 1989) framework of communication from the philosophy of language literature to construct an account of the conversational dynamics which underpin communicative exchanges generally. I then used this to construct an account of the interpersonal relation at play in cases of knowledge transmission specifically. This, in turn, provided us with a fully developed view of knowledge transmission – a view of knowledge transmission which is consistent with anti-reductionism as it is traditionally construed – in light of which I constructed my theoretical elaboration of our candidate explicatum. In the course of constructing my theoretical elaboration of our candidate explicatum, I solved three immediate problems for Grice's framework of communication (i.e., the inaccurateness problem, the incompleteness problem, and the obscureness problem) as well as addressed items (1) through (10) on the substantive list:

1) How assertion puts a hearer in a position to know the truth of what is asserted.
2) The conditions under which assertion puts a hearer in a position to know the truth of what is asserted.
3) Speaker sincerity and insincerity (i.e., the sense in which assertion relates to speakers' beliefs).
4) Speaker reliability and unreliability (i.e., the sense in which sincere assertion relates to the reliability of speakers' beliefs).
5) Expressing belief (i.e., the sense in which assertion expresses or manifests speakers' beliefs).
6) Representing oneself as knowledgeable (i.e., the sense in which speakers are typically thought to implicitly represent themselves has having knowledge in what they assert).
7) Commitment (i.e., the sense in which a speaker confers a commitment upon themselves to the truth of the proposition they assert).

8) Responsibility (i.e., the sense in which speakers incur a responsibility for the belief that the hearer is intended to acquire).
9) Challenges (i.e., the sense in which assertions can be contested by querying the speaker's epistemic standing on the proposition they have asserted).
10) Retractions (i.e., the sense in which a speaker can retract an assertion when they regard themselves as no longer in a position to believe or vindicate what they have asserted).

In *Exploring the Exotica*, I defended my theoretical elaboration of our candidate explicatum – my Gricean account of assertion – against some potential objections. Here, we learned an important lesson, namely, arguing by theory is an advantageous way of doing philosophy whereas arguing by counterexample is a risky business. In addition, I addressed the final item on the substantive list:

11) Moore's paradox (i.e., the apparent absurdity in attempting to assert '*p* but I do not believe that *p*').

Now, as I have said, my theoretical elaboration of our candidate explicatum not only affords us a vocabulary to help us explain the core phenomena of assertion but also there is no immediate reason to think that it does not afford us a vocabulary to help us explain any outstanding philosophically interesting phenomena peripheral in the study of the speech act as well. These include truth, semantics (i.e., the specification of meanings in linguistic expressions within a context), meta-semantics (i.e., the determination of semantic facts), implicature, presupposition, lying, so-called 'epistemic normativity', knowledge ascriptions, expert disagreement, testimonial injustice, radical interpretation, etc. Indeed, with a new way of talking when we talk about assertion, we are likely to find that there are new ways of understanding such phenomena – we have already found this to be the case in this thesis, in fact. There are, however, some potential lines of future research which I find particularly intriguing.

One is the role of assertion in mass communication, online or otherwise, and especially in political speech. For example, when Donald Trump makes a briefing, one complaint which is often heard is that he is attempting to tell people things which are commonly known to be false. Alternatively, however, we might understand him as making genuine assertions to certain targeted sub-groups of the population in which such common knowledge is not instantiated. Those who make complaints of the sort just mentioned above should therefore be understood as separate from his intended audience, even though they take themselves to be part of that audience. Relatedly, there are the ways in which speakers within the political domain often use jargon when disseminating information. Consider, for example, Theresa May's obfuscated assertion ''Brexit' means Brexit'. Such speakers are so unclear in their meaning that their audiences are not able to respond appropriately and so not able to participate properly in the debate, which, in turn, inhibits political choice.

Another is what we might call 'conversational destabilisation' where, say, persistent sceptics introduce certain sceptical possibilities into the context of conversation illegitimately, which thus frustrates the transmission of knowledge. This, in turn, may result in what we might call 'epistemic destabilisation', where the victim gradually loses confidence in their own epistemic faculties, and perhaps becomes abnormally epistemically dependant on others, which may make them more susceptible to machinations. Interesting case studies might include gaslighters and conspiracy theorists, especially climate-change deniers and anti-vaxxers.

Although I think these potential lines of future research are philosophically interesting in and of themselves, they seem particularly pertinent given the ongoing coronavirus pandemic.

References

Anscombe, G. E. M. (1957) *Intention*. Harvard. Harvard University Press.

Anscombe, G.E.M. (1979) What is it to believe someone? in *Rationality and religious belief* (ed.) C.F. Delaney: Notre Dame (pp. 141–151).

Armstrong, D. (1973) *Belief, truth, and knowledge*. Cambridge: Cambridge University Press.

Audi, R. (2006) 'Testimony, credulity, and veracity' in *The epistemology of testimony* (eds.) J. Lackey & E. Sosa. Oxford: Oxford University Press (pp. 25–49).

Austin, J. L. (1962) *How to do things with words*. Cambridge: Harvard University Press.

Bach, K. & Harnish, R. M. (1979) *Linguistic communication and speech acts*. Massachusetts: MIT Press.

Baldwin, T. (2003) *G.E. Moore: selected writings*. Routledge.

Barney, R. (2016) '[Aristotle], on trolling', *Journal of the American Philosophical Association* 2(2): 193-195.

Black, M. (1952) 'Saying and disbelieving', *Analysis* 13(2): 25–33.

Blome-Tilman, M. (2014) *Knowledge and presuppositions*. Oxford: Oxford University Press.

Bourne, C., Caddick Bourne, E., & Jarmy, C. (2016) 'The basis of correctness in the religious studies classroom', *Journal of Philosophy of Education* 50: 669–688.

Bourne, C. & Caddick Bourne, E. (2018) 'Conversational perversions, implicature and sham cancelling in *Othello*' in *The Routledge companion to Shakespeare and philosophy* (eds.) C. Bourne & E. Caddick Bourne. Routledge (pp. 146–160).

Brandom, R. (1983) 'Asserting', *Noûs* 17(4): 637-650.

Brandom, R. (1994) *Making it explicit: reasoning, representing, and discursive commitment.* Harvard: Harvard University Press.

Brandom, R. (2001) 'Modality, normativity, and intentionality', *Philosophy and Phenomenological Research 63*(3): 611–623.

Brun, G. (2016) 'Explication as a method of conceptual re-engineering', *Erkenntnis* 81(6): 1211–1241.

Burgess, A., Cappelen, H., & Plunkett, D. (eds.) (2020) *Conceptual ethics and conceptual engineering*. Oxford: Oxford University Press.

Burge, T. (1993) 'Content preservation', *The Philosophical Review* 102(4): 457-488.

Burge, T. (2013) 'Postscript: content preservation' in *Cognition through understanding: self-knowledge, interlocution, reasoning, reflection: philosophical essays*, *volume 3* (ed.) T. Burge. Oxford: Oxford University Press.

Cappelen, H. (2011) 'Against assertion' in *Assertion: new philosophical essays* (eds.) J. Brown & H. Cappelen. Oxford: Oxford University Press (pp. 21–47).

Cappelen, H. (2020a) 'Assertion: a defective theoretical category' in *The Oxford handbook of assertion* (ed.) S. Goldberg. Oxford: Oxford University Press (p. 139-155).

Cappelen, H. (2020b) 'Conceptual engineering: the master argument' in *Conceptual ethics and conceptual engineering* (eds.) A. Burgess, H. Cappelen & D. Plunkett. Oxford: Oxford University Press (pp. 132–151)

Carnap, R. (1947) *Meaning and necessity*. Chicago: University of Chicago Press.

Carnap, R. (1950) *Logical foundations of probability*. Chicago: University of Chicago Press.

Carson, T. L. (2006) 'The definition of lying', *Noûs* 40(2): 284–306.

Cassam, Q. (2010) 'Judging, believing and thinking', *Philosophical Issues* 20: 80-95.

Chalmers, D. (2011) 'Verbal disputes', *The Philosophical Review* 120(4): 515-566.

Chan, T. & Kahane, G. (2011) 'The trouble with being sincere', Canadian Journal of Philosophy 41(2): 215–34.

Chisholm, R. & Feehan, T. (1977) 'The Intent to Deceive', *The Journal of Philosophy* 74: 143-159.

Clark, A. & Chalmers, D. (1998) 'The extended mind', *Analysis* 58(1): 7–19.

Coady, C. (1992) *Testimony: a philosophical study*. Oxford: Oxford University Press.

Comesaña, J. (2010) 'Evidentialist reliabilism', Noûs 94: 571–601.

Davidson (1973) 'Radical interpretation', *Dialectica* 27: 314–28

Davidson, D. (1984) *Inquiries into truth and interpretation*. Oxford: Oxford University Press.

Davidson, D. (1991) 'Three varieties of knowledge' in *A.J. Ayer memorial essays: Royal Institute of Philosophy supplement*, 30 (ed.) A. Phillips Griffiths. Cambridge: Cambridge University Press (pp. 153–166).

Davidson, D. (2001) *Subjective, intersubjective, objective*. Oxford: Oxford University Press.

Deutsch, M. (2020) 'Speaker's reference, stipulation, and a dilemma for conceptual engineers', *Philosophical Studies* 177(12): 3935–3957.

Dummett, M (1994) 'Testimony and Memory' in *Knowing from words* (eds.) B. K. Matilal & A. Chakrabarti. Dordrecht: Kluwer Academic Publishers (pp. 251–272).

Dutilh Novaes, C. & Reck, E. (2017) 'Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization', *Synthese* 194(1): 195–215.

Fallis, D. (2009) 'What is lying?', *The Journal of Philosophy* 106(1): 29–56.

Fallis, D. (2012) 'Lying as a violation of Grice's first maxim of quality', *Dialectica* 66(4): 563–581.

Fallis, D. (2014) 'Are bald-faced lies deceptive after all?', *Ratio* 27(3): 81–96.

Faulkner, P. (2011) *Knowledge on trust*. Oxford: Oxford University Press.

Faulkner, P. (2018) 'Collective testimony and collective knowledge', *Ergo* 5: 103–126.

Fricker, E. (1987) 'The epistemology of testimony', *Proceedings of the Aristotelian Society* 61: 57-83.

Fricker, E. (2012) 'Stating and insinuating', *Aristotelian Society Supplementary*, Volume 86 (1):61-94.

Fricker, E. (2015) 'How to make invidious distinctions amongst reliable testifiers', *Episteme* 12 (2): 173-202.

Fricker, M. (2012) 'Group testimony? The making of a collective good informant', *Philosophy and Phenomenological Research* 84(2):249-276.

Garcia-Carpintero, M. (2018) 'Sneaky assertions', *Philosophical Perspectives* 32(1): 188-218.

Gelfert, A. (2014) *A critical introduction to testimony*. Bloomsbury: London.

Goldberg, S. (2005) 'Testimonial knowledge through unsafe testimony', *Analysis* 65(4):302-311.

Goldberg, S. (2010) 'Assertion, testimony, and the epistemic significance of speech', *Logos and Episteme* 1(1):59-65 (2010)

Goldberg, S. (2015) *Assertion: on the philosophical significance of assertoric speech*. Oxford: Oxford University Press.

Goldberg, S. (2018) *To the best of our knowledge: social expectations and epistemic normativity*. Oxford: Oxford University Press.

Goldman, A. (1979) 'What Is justified belief?' in *Justification and knowledge* (ed.) G.S. Pappas. Dordrecht: Reidel (pp. 1–25)

Goldman, A. (1986) *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. (2011) 'Toward a synthesis of reliabilism and evidentialism' in *Evidentialism and its discontents* (ed.) T. Dougherty. New York: Oxford University Press (pp. 123–151).

Grandy, R. (1973) 'Reference, meaning, and belief', *The Journal of Philosophy* 70(14): 439–452.

Greco, J. (2012) 'Recent work on testimonial knowledge', *American Philosophical Quarterly* 49(1):15-28.

Grice, H. P. (1947–1989) 'H. Paul Grice papers', *banc mss 90/135c*, The Bancroft Library, University of California, Berkeley.

Grice, H.P. (1957) 'Meaning', *The Philosophical Review* 66: 377–88.

Grice, H.P. (1969) 'Utterer's meaning and intention', *The Philosophical Review* 78(2): 147–177

Grice, H. P. (1975) 'Logic and conversation' in *The logic of grammar* (eds.) D. Davidson & G. Harman. Encino, CA: Dickenson (pp. 64–75).

Grice, H. P. (1989) *Studies in the way of words*. Harvard: Harvard University Press.

Groenendijk, J. & Stokhof, M. (1997) 'Questions', in *Handbook of logic and language* (eds.) J. van Benthem & A. ter Muelen. Cambridge: MIT Press (pp. 1055–1124).

Hamblin, C. L. (1973) 'Questions in Montague English', *Foundations of language* 10(1): 41–53.

Haslanger, S. (2000) 'Gender and race: (what) are they? (What) do we want them to be?', *Noûs* 34(1): 31-55.

Hawley, K. (2014) 'Trust, distrust and commitment', *Noûs* 48(1):1-20.

Hinchman, E. S. (2005) 'Telling as inviting to trust', *Philosophy and Phenomenological Research*, 70(3): 562-587.

Hinchman, E. S. (2020) 'Assertion and testimony', in *The Oxford handbook of assertion* (ed.) S. Goldberg. Oxford: Oxford University Press (pp. 555-579).

Hobbes, T. (1651) *Leviathan, with selected variants from the Latin edition of 1668* (ed.) E. Curley (1994). Indianapolis: Hackett.

Hume, D. (1748) *An enquiry concerning human understanding.* Indianapolis: Hackett Publishing Company.

Hughes, J. (1984) 'Group speech acts', *Linguistics and Philosophy* 7(4):379 - 395.

Ichikawa, J.J. (2017) *Contextualizing knowledge*. Oxford: Oxford University Press.

International Astronomical Union. (2006a) The IAU draft definition of 'planet' and plutons' (Press Release iau0601). Accessed 10 December 2020, from https://www.iau.org/news/pressreleases/detail/iau0601/.

International Astronomical Union. (2006b) Resolution B5: Definition of a planet in the solar system. Accessed 10 December 2020, from https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf.

Karttunen, L. (1977) 'Syntax and semantics of questions', *Linguistics and Philosophy* 1(1):3-44.

Keiser, J. (2016) 'Bald-faced lies: how to make a move in a language game without making a move in a conversation', *Philosophical Studies* 173(2): 461-477.

Keiser, J. (2017) *Languages and Language Use*. Unpublished doctoral dissertation. Yale University, New Haven, Connecticut.

Kitcher, P. (2008) 'Carnap and the caterpillar', *Philosophical Topics* 36(1): 111–127.

Kripke, S. (1975) 'Outline of a theory of truth', *The Journal of Philosophy* 72(19): 690–716.

Lackey, J. (2008) Learning from words: testimony as a source of knowledge. Oxford: Oxford University Press.

Lackey, J. (2015) 'A deflationary account of group testimony' in Essays in collective epistemology (ed.) J. Lackey. Oxford: Oxford University Press (pp. 69–94).

Lackey, J. (2018) 'Group assertion', *Erkenntnis* 83(1): 21-42.

Leslie, S. (2017) 'The original sin of cognition: fear, prejudice, and generalization', *The Journal of Philosophy 114: 393–421.*

Lewis, D. (1969) *Conventions.* Cambridge: Harvard University Press.

Lewis, D. (1974) 'Radical Interpretation', *Synthese* 23: 331–344.

Lewis, D. (1975) 'Languages and language' in *Minnesota studies in the philosophy of science*, Volume VII (ed.) K. Gunderson. Minneapolis: University of Minnesota Press (pp. 3–35).

Lewis, D. (1980) 'A subjectivist guide to objective chance' in *Studies in inductive logic and probability*, Volume II. (eds.) R. C. Jeffrey. Berkeley: University of California Press (pp. 263-293).

Lewis, D. (1994) 'Humean supervenience debugged', *Mind* 103: 473–490.

Lewis, D. (1996) 'Elusive knowledge', *Australasian Journal of Philosophy* 74(4): 549-567.

Lipton, P (2007) 'Alien abduction: inference to the best explanation and the management of testimony', *Episteme* 4(3):238-251.

MacFarlane, J. (2011) 'What is assertion?' in *Assertion: new philosophical essays* (eds.) J. Brown & H. Cappelen. Oxford: Oxford University Press (pp. 79–96).

Mahon, J. E. (2014) 'History of deception: 1950 to the present', *Encyclopedia of Deception*, New York: Sage: 618–619.

Maitra, I. (2018) 'Lying, acting, and asserting' in *Lying: language, knowledge, ethics, and politics* (ed.) A. Stokke & E. Michaelson. Oxford: Oxford University Press (pp. 65–83).

McMyler, B. (2011) *Testimony, trust, and authority*. Oxford: Oxford University Press.

Meibauer, J. (2014) 'Bald-faced lies as acts of verbal aggression', *Journal of Language Aggression and Conflict* 2(1): 127–150.

Meijers, A. (2007) 'Collective speech acts' in *Intentional acts and institutional facts: essays on John Searle's social ontology* (ed.) S. L. Tsohatzidis. Springer, Dordrecht (pp. 93–110).

Mellor, D. H. (1988) 'Inaugural lecture: the warrant of induction' in *Matters of metaphysics*. Cambridge: Cambridge University Press (pp. 254–268).

Milić, I. (2017) 'Against selfless assertions', *Philosophical Studies* 174(9): 2277-2295.

Montminy, M. (2020) 'Testing for assertion' in *The Oxford handbook of assertion* (ed.) S. Goldberg. Oxford: Oxford University Press (pp. 371–389).

Moran, R. (2005) 'Getting told and being believed', *Philosophers' Imprint* 5(5): 1-29.

Moran, R. (2018) *The exchange of words*. Oxford: Oxford University Press.

Nagel, T. (1969) 'Sexual perversion', *The Journal of Philosophy* 66 (1):5-17.

Neale, S. (1992) 'Paul Grice and the philosophy of language', *Linguistics and Philosophy* 15(5): 509-559.

Owens, D. (2006) 'Testimony and assertion', *Philosophical Studies* 130: 105-129.

Papineau, D. (2013) 'There are no norms of belief' in *The aim of belief* (ed.) T. Chan. Oxford: Oxford University Press (pp. 64–80).

Peet, A (2015) 'Testimony, pragmatics, and plausible deniability', *Episteme* 12(1):29-51.

Peirce, C. S. (1934) *Belief and judgment*. Cambridge: Harvard University Press.

Pinder, M. (2019) 'Conceptual engineering, metasemantic externalism and speaker-meaning', *Mind*. Early online view.

Pinder, M. (2020a) 'What ought a fruitful explicatum to be?', *Erkenntnis*: 1–20.

Pinder, M. (2020b) 'On Strawson's critique of explication as a method in philosophy', *Synthese* 197(3): 955-981.

Plunkett, D. & Sundell, T. (2013) 'Disagreement and the semantics of normative and evaluative terms', *Philosophers' Imprint* 13(23):1–37.

Pollock, J. L. (1987) 'Defeasible reasoning', *Cognitive Science* 11: 481–518.

Railton, P. (1994) 'Truth, reason, and the regulation of belief', *Philosophical Issues* 5: 71–93.

Reid, T. (1764) *An inquiry into the human mind on the principles of common sense*. Indianapolis: Hackett Publishing Company.

Roberts, C. (1996) 'Information structure in discourse: towards an integrated formal theory of pragmatics', *OSU Working Papers in Linguistics*, 49: 91–136.

Rumfitt, I. (1995) 'Truth Conditions and Communication', *Mind* 104(416): 827–862.

Saul, J. M. (2012) *Lying, misleading, and what is said: an exploration in philosophy of language and in ethics*. Oxford: Oxford University Press.

Searle, J. (1969) *Speech acts: an essay in the philosophy of language*. Oxford: Oxford University Press.

Scharp, K. (2013). *Replacing truth*. Oxford University Press.

Schelling, T. (1960) *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schiffer, S. (1972) *Meaning*. Oxford: Clarendon Press.

Schoubye, A. & Stokke, A. (2016) 'What is said?', *Noûs* 50(4): 759-793.

Schupbach, J. (2017) 'Experimental explication', *Philosophy and Phenomenological Research* 94(3): 672–710.

Simion, M. (2017) 'The 'should' in conceptual engineering', *Inquiry* 61: 914–928.

Simion, M. (2019) 'Conceptual engineering for epistemic norms', *Inquiry.* Early online view.

Simion, M. (2020) 'Testimonial contractarianism', *Noûs.* Early online view.

Simion, M. & Kelp, C. (2020) 'Conceptual innovation', *Noûs* 54(4): 985-1002.

Sorensen, R. (2007) 'Bald-faced lies! Lying without the intent to deceive', *Pacific Philosophical Quarterly* 88: 251–264.

Sosa, E. (1994) 'Testimony and coherence' in *Knowing from words* (eds.) B. K. Matilal & A. Chakrabarti. Dordrecht: Kluwer Academic Publishers (pp. 59-67).

Soter, S. (2006) 'What is a planet?', *Astronomical Journal* 132(6): 2513–2519.

Sperber, D. & Wilson, D. (1986) *Relevance: communication and cognition*. London: Blackwell.

Sperber, D. & Wilson, D. (2002) 'Truthfulness and relevance', *Mind* 111(443): 583-632.

Stalnaker, R. (1978) 'Assertion', *Syntax and Semantics* 9: 315-332.

Stalnaker, R. (1999) *Context and content*. Oxford: Oxford University Press.

Stalnaker, R. (2002) 'Common ground', *Linguistics & Philosophy* 25: 701–21.

Stern, S. A. & Levison, H. (2002) 'Regarding the criteria for planethood and proposed planetary classifcation schemes', *Highlights of Astronomy* 12: 205–213.

Stevenson, L. (1993) 'Why believe what people say?', *Synthese* 94: 429–451.

Stokke, A. (2013) 'Lying and asserting', *The Journal of Philosophy* 110(1): 33–60.

Stokke, A. (2016) 'Truthfulness and Gricean cooperation', *Grazer Philosophische Studien* 93(3): 489-510.

Stokke, A. (2018) *Lying and insincerity*. Oxford, UK: Oxford University Press.

Strawson, P. F. (1963) 'Carnap's views on constructed systems versus natural languages in analytic philosophy' in *The philosophy of Rudolf Carnap* (ed.) P. Schilpp. Lasalle, IL: Open Court (pp. 503–518).

Strawson, P.F. (1964) 'Intention and convention in speech acts', *The Philosophical Review* 73(4): 439–460.

Sullivan, N (2012) 'Don't feed the trolls', *Fluent Conference*. Accessed 15 December 2020 from https://www.youtube.com/watch?feature=player_embedded&v=ulNSlES1Fds

Sullivan, A. (2019) 'Donald Trump and the art of the lie', *Intelligencer*: New York Magazine. Accessed 15 December 2020 from https://nymag.com/intelligencer/2019/06/andrew-sullivan-donald-trump-and-the-art-of-the-lie.html

Thalberg, I. (1962) 'Intending the impossible', *Australasian Journal of Philosophy* 40(1): 49-56.

Thomasson, A. (2020) 'A pragmatic method for normative conceptual work', in A. Burgess, H. Cappelen, & D. Plunkett (eds.), *Conceptual ethics and conceptual engineering*. Oxford University Press (pp. 435– 459).

Tollefsen, D. (2007) 'Group testimony', *Social Epistemology* 21(3):299 – 311.

Tollefsen, D. (2009) 'WIKIPEDIA and the epistemology of testimony', *Episteme* 6(1):8-24.

Tollefsen, D. (2011) 'Groups as rational sources' in *Collective epistemology* (eds.) H.B Schmid, D. Sirtes, & M. Weber. De Gruyter (pp. 11–22).

Tollefsen, D. (2020), 'Can groups assert that p?' in *The Oxford Handbook of Assertion* (eds.) S. Goldberg (pp. 327–344).

Turri, J. (2010) 'Prompting challenges', *Analysis* 70: 456–462.

van Inwagen, P. (2008) 'How to think about the problem of free will', *Journal of Ethics* 12(3/4): 327–41.

Velleman, D. (2000) 'On the aim of belief' in *The possibility of practical reason (ed.)* D. Velleman. New York: Oxford University Press (pp. 244–281).

Viebahn, E. (2017) 'Non-literal lies', *Erkenntnis* 82: 1367-1380.

Wedgwood, R. (2002) 'The aim of belief', *Philosophical Perspectives* 16(s16): 267–97.

Wedgwood, R. (2007) *The nature of normativity*. New York: Oxford University Press.

Welbourne, M. (1986) *Community of knowledge*. Aberdeen: Aberdeen University Press.

Whiting, D. (2010) 'Should I believe the truth?', *Dialectica*, 64(2): 213–224.

Whiting, D. (2013) 'Stick to the facts: on the norms of assertion', *Erkenntnis* 78(4): 847-867.

Williams, B. (1973) 'Deciding to believe' in *Problems of the self* (ed.) B. Williams. Cambridge, MA: Cambridge University Press (pp. 136–151).

Williamson, T. (1996) 'Knowing and asserting', Philosophical Review 105 (4):489-523.

Williamson, T. (2000) *Knowledge and its limits*. Oxford: Oxford University Press.

Williamson, T. (2007) *The philosophy of philosophy*. London: Blackwell.

Wittgenstein, L. (1953) *Philosophical investigations*. (eds.) G.E.M. Anscombe & R. Rhees (trans.) G.E.M. Anscombe. Oxford: Blackwell.

Wright, S. (2018) *Knowledge transmission*. Routledge.